Phylogenetic Analysis of Type II Toxin-Antitoxin Systems

Camps Lab
Nitya Jain

## I. Abstract

Recent studies have highlighted the incredible specificity of toxin-antitoxin type II (TA) systems in bacteria. These toxic pairs have been implicated in virulence and plasmid maintenance. TA systems consist of a toxin protein and a corresponding antitoxin protein that binds to the toxin to inhibit it. The toxin targets essential bacterial machinery such as growth and replication factors. Due to their toxic nature, they are highly specific pairs where the antitoxin cannot act on another of a similar species. Emerging evidence suggests that multiple antitoxin species can bind to the same toxin. This challenges the traditional 1:1 genetic specificity relationship of TA systems therefore prompting a re-evaluation of TA system classification, specifically in Type II TA systems. This paper will present findings from phylogenetic analyses to understand this phenomenon of multiple antitoxin families binding to the same toxin family and its implications for classifying Type II TA systems.

## II. Introduction

Plasmids are self-replicating extrachromosomal DNA, that carry nonessential genes. These plasmids are widely distributed across the prokaryotic kingdom[2]. Certain genetic elements of plasmids can provide bacteria with useful functions such as virulence and drug resistance, allowing long-term survival of the species. These plasmids have a metabolic cost to their host and thus need maintenance genes to ensure their presence in a bacterial colony. One such maintenance gene is type II TA systems. TA systems were first discovered on plasmids, where they discovered the post-segregational killing (PSK) model[3]. PSK occurs during bacterial replication. Type II TA systems for a toxin-antitoxin complex which remaions dormant. The antitoxin is less stable than the toxin, thus once the antitoxin degrades, the toxin becomes active. This mechanism becomes deadly when a bacteria replicates into two daughter cells. If a daughter cell fails to receive a plasmid that contains a TA system gene cassette to continually produce antitoxins. The dormant bound toxins become active; killing the daughter cell. This phenomenon is broadly known as “plasmid addiction.”

However, little is known about the origin of TA systems, especially their evolutionary relationships and functional similarities. Combined with the small size and large divergence among TA system sequences within families; it is likely that multiple types need to be classified or annotated correctly[2]. The aim of previous phylogenetic analyses was to find new TA systems that had yet to be annotated in NCBI. Comparing the TA family-specific trees against each other suggests that they share a common ancestor, especially considering relative structural similarity. However, we can group toxins not only by their structure but their targets. For example, toxins

from the RelE superfamily and ParE superfamily demonstrate that the two families are homologous but are "thought to exert their toxic activity on different targets," indicating points of divergence. Another theory is that TA systems have evolved several times independently. Moreover, current data shows that there are multiple antitoxins paired with a given toxin, breaking away from the specificity model that there is only one antitoxin that will bind to the toxin.

Furthermore, it is unlikely that PSK is their only function. TA systems are found in large numbers with diverse families within host genomes, suggesting other biological functions. Even so, multiple TA systems are found on plasmids, and within the dataset, we found that they are from the same family. Typically, duplicated TA systems are not found within the same genome. Understanding this genetic anomaly will add to our understanding of the plasmids that spread antibiotic resistance and virulence.

## III. Methods

### A. Data

The data was downloaded from NCBI in March of 2023. We downloaded ~3,400 complete genomes of *E.Coli* and 8209 plasmids from these complete genomes.

### B. Computational Methods

Filtering methodology:

We first filtered the dataset for all duplicated plasmids and TA systems using the Pandas library in python. We then filtered out entries containing multiple antitoxin hits for one toxin to reduce noise in generating the alignments and trees.

MUSCLE:

MUSCLE is a sequence alignment software for protein sequences. We utilized MUSCLE to create multi-sequence alignments for each TA family, split between the toxins and antitoxins. Using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) algorithm, we then utilized MUSCLE to generate distance-based trees. Distance-based guide trees are advantageous for larger datasets as they are more computationally efficient and scalable, compared to other methods.

ITOL:

The Interactive Tree of Life (ITOL) is an online tool to display and annotate phylogenetic trees. We utilized this tool to visualize the trees from MUSCLE and to annotate the subgroups.

The GitHub repository contains all of the code and input files provided in the Supplementary Information section.

**Results**

Out of the 11 TA families in the dataset, we took a closer look at the CcdA/CcdB (CcdAB) family as this family has been well-documented because it is a Type II TA system and densely collected.

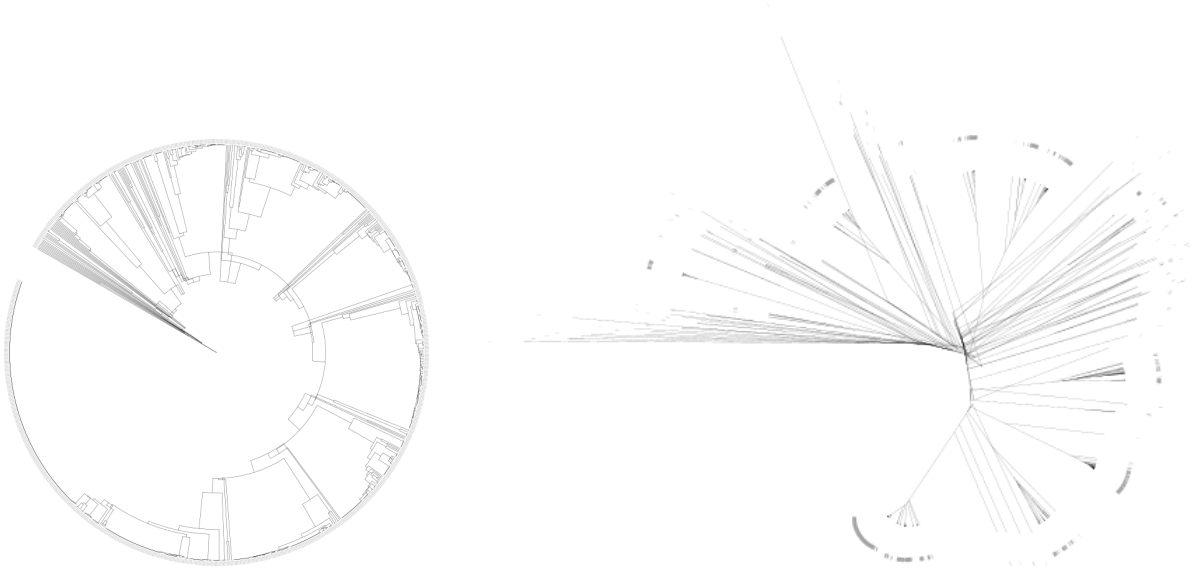

Fig1: Unrooted representation of CcdB toxin (left-hand side) and CcdA antitoxin (right-hand side)

Through the visual comparisons, it can be seen that the antitoxins are less uniform and more dispersed than the toxins. More concrete subgroups/families can be easily differentiated in terms of the toxins compared to the antitoxins. Moreover, there appear to be more subgroups from the antitoxins that are closely related to each other compared to the toxins by analyzing the number of branches coming out of the main descendant node.

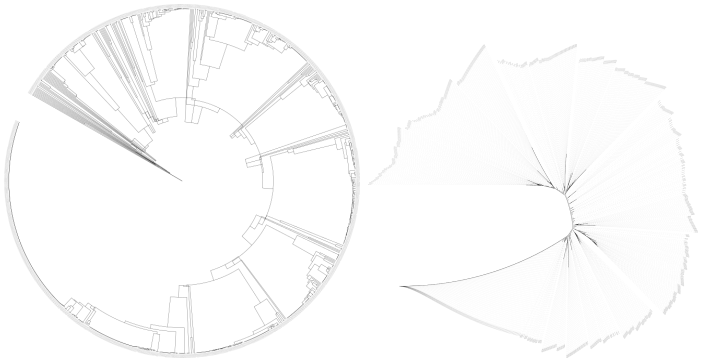

Fig2: Unrooted representation of CcdB toxin (left-hand side) and CcdA antitoxin (right-hand side), positive branch lengths only

We removed the negative branch lengths by setting them to 0 to reduce noise in the figures. The toxin remains largely unchanged, whereas the antitoxin sees a vast difference. The antitoxin subgroups are still widely spread out compared to the toxins but have become more groupable in this way.

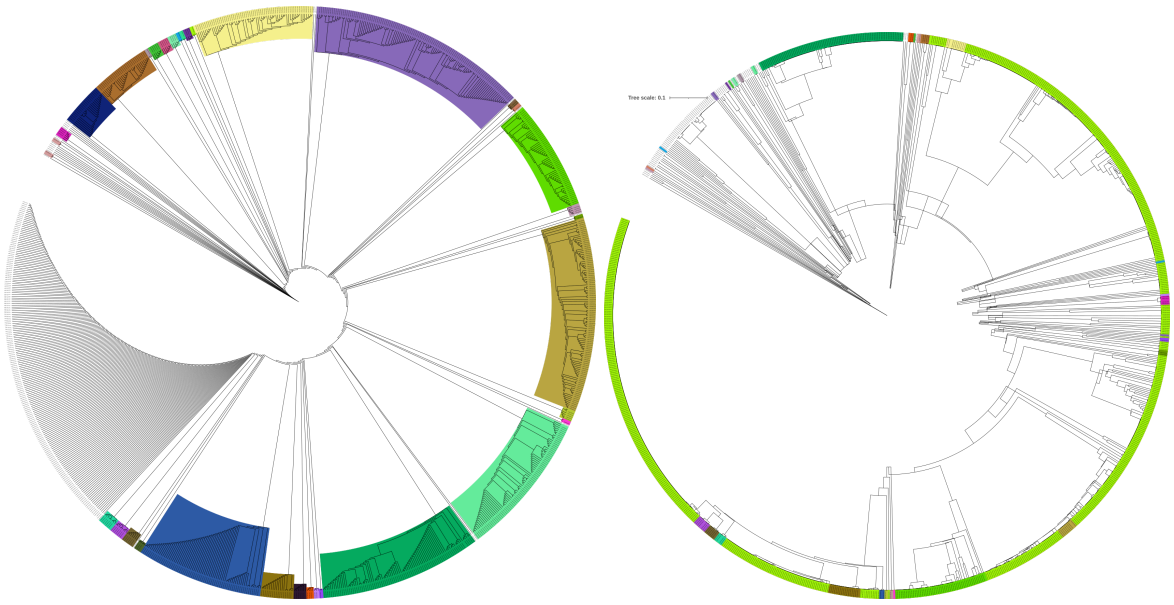

Fig3: Circular representation of CcdB toxin (left-hand side) and CcdA antitoxin (right-hand side) color-coded by groups

We systematically grouped clades containing two or more leaves for the toxins to assess the consistency of groupings. Initially, clade groupings were established based on the toxins, and subsequently, these groupings were applied to the corresponding antitoxins. Consequently, the toxin groupings yielded more distinct patterns. In analyzing the antitoxin group, a noticeable disproportion in representation was observed, characterized by the predominance of a specific

subgroup, prominently highlighted in lime green. This dominant subgroup was accompanied by smaller subgroups nested within it.

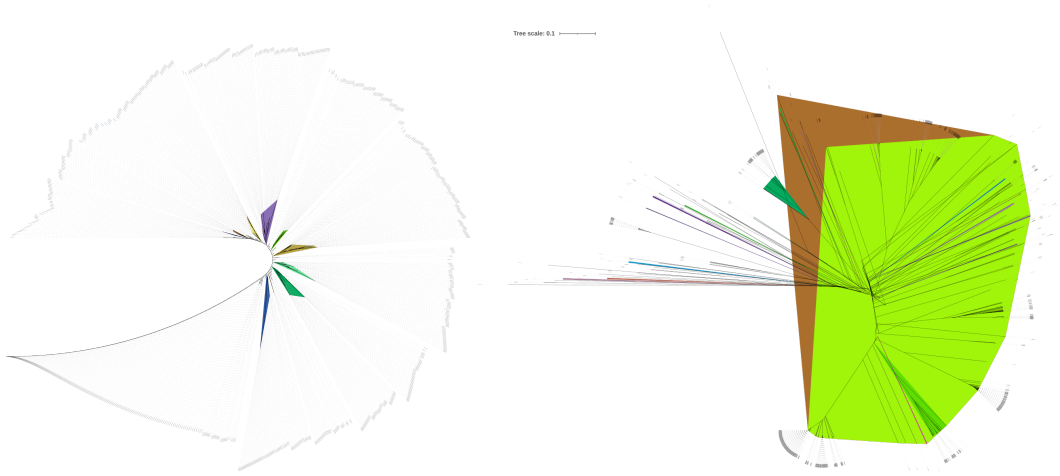

Fig4: Unrooted representation of CcdB toxin (left-hand side) and CcdA antitoxin (right-hand side)

The unrooted comparison of the toxin and antitoxin depicts a stark contrast in the distribution patterns. Given the close relationship between these groups in the antitoxin tree, certain smaller subgroups may be obscured, particularly beneath the dominant lime green subgroup. This is not seen in the toxin as the subgroups are fairly distributed.

**Discussion**

Our main hypothesis was to show how different antitoxins are from their respective toxins, as antitoxins would be more closely related to one another and have more dichotomy compared to the toxins. The data filtration reduces the noise level in the phylogenetic trees as we are looking at TA systems that are inherently closely related to one another. Constructing the phylogenies of the toxins and the antitoxins separately allows us to analyze their comparative evolutionary relationships. The null assumption is that the phylogenies would look similar as they are typically formed as pairs, except for a couple of TA systems with three components. However, the phylogenies show that the antitoxins are less closely related than the toxins, proving the hypothesis. Furthermore, we did not expect to see a toxin that did not maintain specificity as multiple subgroups were found in one subgroup.

Further analysis should be done to support this hypothesis by looking at the correspondence between the toxin and antitoxins to specific *E.coli* genomes or plasmids. Another analysis would be to highlight the subgroups based on conjugated and nonconjugated plasmids. We can also utilize AlphaFold, an artificial intelligence program that performs predictions of protein structures, to see if the toxin and antitoxin structurally fit together. Moreover, we can identify if the antitoxin fits other toxins structurally, which may address any concerns about misclassification.

Lastly, this overall analysis should be performed using the other well-studied TA systems, ParD and PIN. Previous analysis has shown that these systems' phylogenies follow a similar pattern to the CcdB system, but further analysis is needed to determine if the antitoxins hold the same properties.

**Bibliography**


1. Edgar R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340
2. Guglielmini, J., Szpirer, C., & Milinkovitch, M. C. (2008). Automated discovery and phylogenetic analysis of new toxin-antitoxin systems. *BMC microbiology*, *8*, 104. https://doi.org/10.1186/1471-2180-8-104
3. Jurėnas, D., Fraikin, N., Goormaghtigh, F. *et al.* Biology and evolution of bacterial toxin–antitoxin systems. *Nat Rev Microbiol* 20, 335–350 (2022). https://doi.org/10.1038/s41579-021-00661-1
4. Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic acids research*, *49*(W1), W293–W296. https://doi.org/10.1093/nar/gkab301


**Appendix**

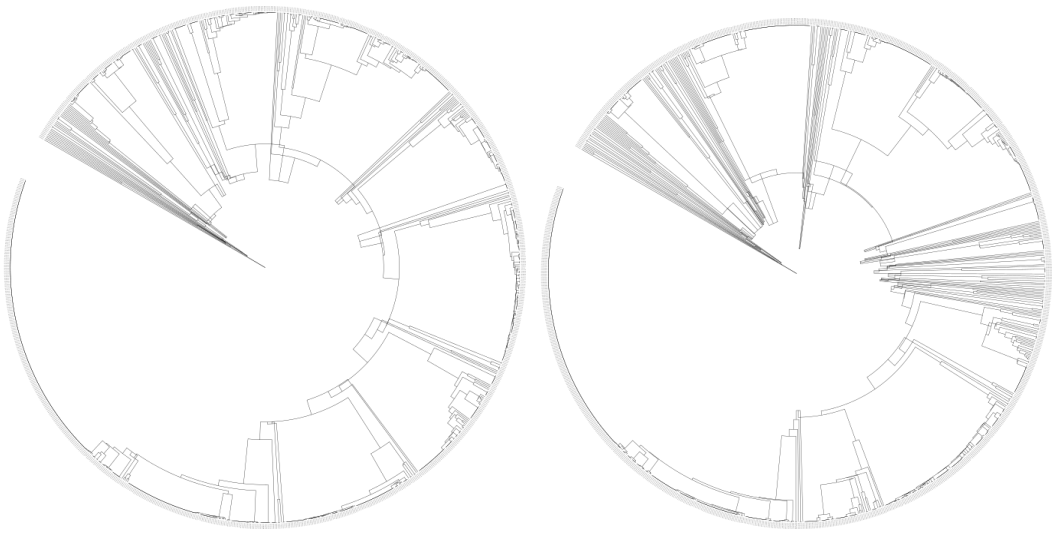

Fig6: Circular representation of PIN toxin (left-hand side) and antitoxin (right hand side)

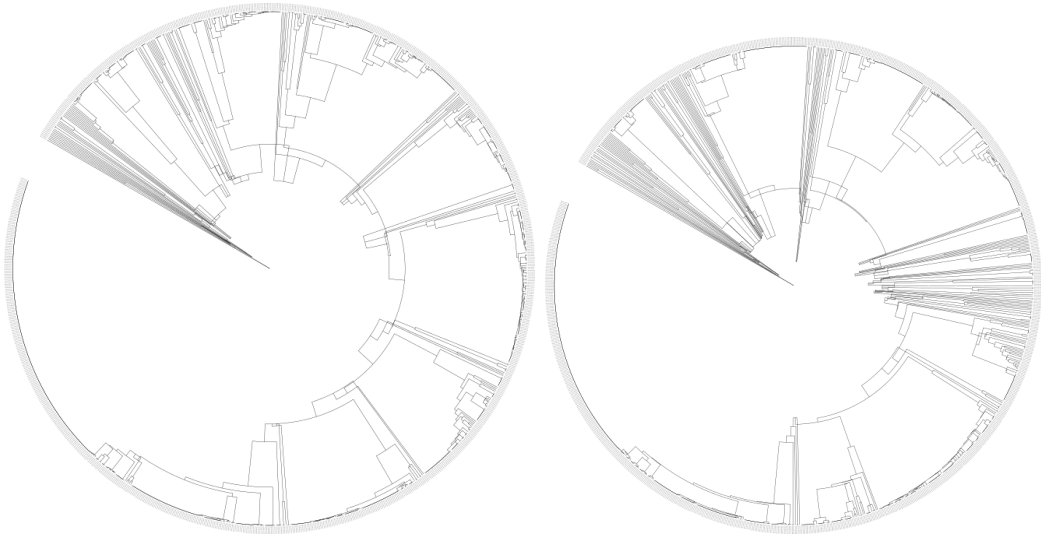

Fig7: Circular representation of ParE toxin (left-hand side) and antitoxin (right hand side)

Access to external files:
https://drive.google.com/drive/folders/1c685AOLrc5Sk2itZb4TPaSekLVJ8oTIB?usp=sharing