Discovery and Annotation of Two Phages that Infect *Microbacterium foliorum*: Tedro and BAjuniper

Grace Anderson, Olivia Anderson, Clayton Bosma, Erin Brouwer, Kleyton DeGroot, Owen Hede, Daiki Jonouchi, Dylan Kirkeby, Jayron Klinghagen, Kate Kralik, Julia Kutz, Olivia Lott, Victoria Pavik, Cayli Penner, Leah Rozeboom, Jett Skrien, Jessica Slight, Maegan Stokes, Jonas Tiensvold, Kyra Wajer, Alaena Trevino, Byron Noordewier, and Sara S. Tolsma
Department of Biology, Northwestern College, Orange City, IA 51041

Abstract

We isolated and purified Tedro and BAjuniper which infect the host *Microbacterium foliorium*. Tedro is a lytic, cluster EF phage isolated from soil collected in Hawarden, Iowa. Its genome is 56,197 bp long, circularly permuted, and includes 83 protein-coding genes and no tRNA genes. We are examining two of Tedro's genes, genes 56 and 57, both of which are predicted to encode a DnaE-like DNA polymerase III (alpha) in more detail. Tedro_57 is twice as large as Tedro_56 so we are using additional bioinformatic tools to understand these genes. BAjuniper was isolated from soil collected in a garden in Orange City, Iowa. Its genome is 41,985 bp long. It was assigned to cluster EB. BAjuniper's genome includes one tRNA gene and we will finalize BAjuniper's annotation shortly.

Introduction

Bacteriophages (phages), viruses that infect bacteria, are the most abundant infectious particles on Earth, with an estimated 10^{31} different kinds in existence (Strange, et al. 2021). They are found in a variety of environments such as soil, water, and the human body. Phages are useful as model organisms for studying genetic processes. They also have beneficial applications clinically, for example, phage therapy utilizes phages to target and kill harmful bacteria within the human body (Keen 2015). Phages act by injecting their DNA into a host bacterial cell. Once a phage has successfully infected a bacterium, it will undergo one of two methods of reproduction. A lytic phage will cause a bacterium to burst, releasing many more phages into the environment, whereas a lysogenic phage integrates its genome into the bacterial genome to be passed onto daughter bacterial cells (Kasman and Porter 2022).

Northwestern College is part of the SEA-PHAGES program, which involves undergraduate students in phage research. Our goal was to annotate the entire genome of our phages, Tedro and BAJuniper, both of which were discovered by Northwestern College students. Annotation of a phage involves determining the start site of each gene within the genome as well as determining a function, when possible. Computer software is used for autoannotation, but it makes mistakes, so it is important for each phage to be manually annotated as well. During our annotation of Tedro, genes Tedro_56 and Tedro_57 caught our attention. We called both genes DnaE-like DNA polymerase III alpha but because they were very different sizes, we performed further analysis of the gene products using Alphafold. Ultimately, our goal is that our research contributes to a better understanding of phages and further innovative applications of phages in the lab and clinic.

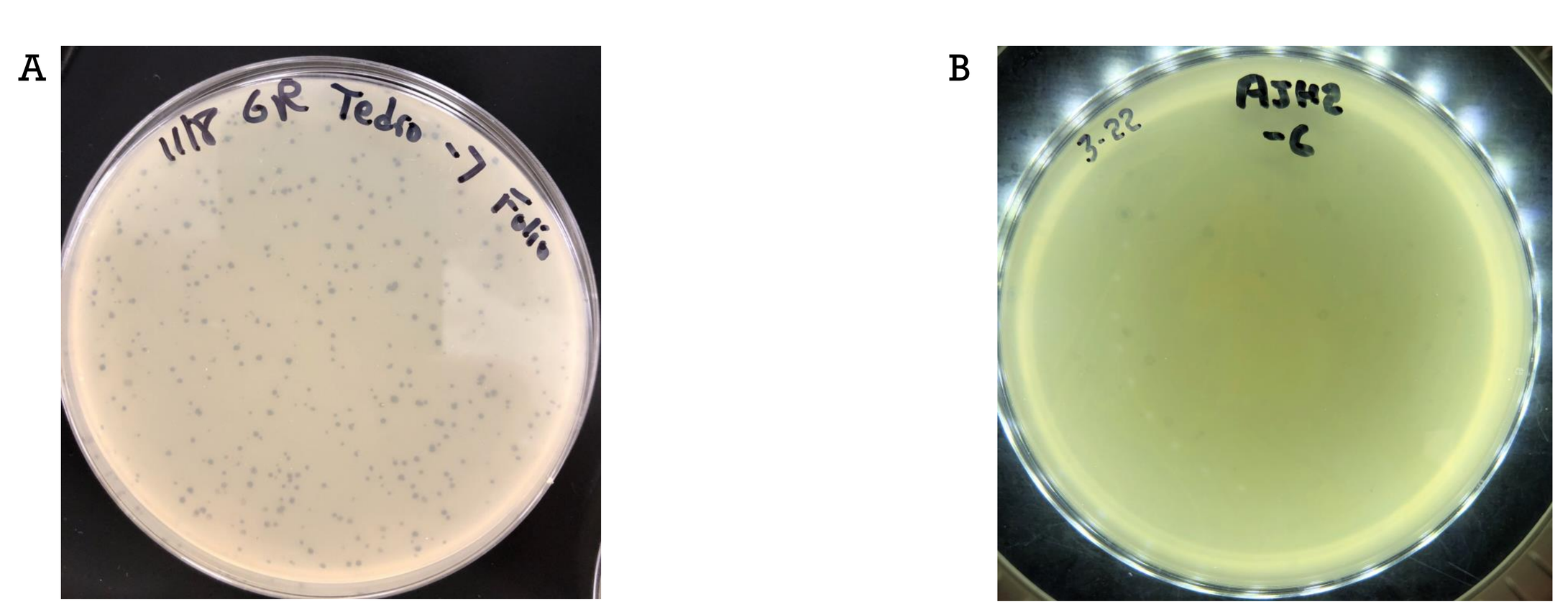


Figure 1. Plaque assays of Tedro and BAjuniper. A. Tedro produces clear, round plaques with a mean diameter of 1.0 mm +/- 0.3 mm. B. BAjuniper produces turbid plaques of variable shape with a mean diameter of 6.1 mm +/- 2.7 mm. Image analysis was performed with ImageJ (Schneider, et al. 2012).

Table 1. Characteristics of Microbacterium phages Tedro and BAjuniper.

	Place of Discovery	Cluster	Cluster Lifecycle	Genome Length and Type	Genome End Type	# of Protein Coding Genes	GC Content
Tedro	Hawarden, IA	EF	Lytic	56,197 bp Circular	circularly permuted	81	63.7 %
BAjuniper	Orange City, IA	EB	Lytic	41,985 bp Linear	3′ sticky overhang	TBD	68.8%

Materials and Methods

Tedro was discovered Erika McKenney and Garrett Raymon in 2022. BAjuniper was discovered by Aimee Hulstein in 2022. Both phages infect *Microbacterium foliorum* NRRL B-24224 and were isolated by direct isolation (Protocol 5.2, https://seaphagesphagediscoveryguide).

Viruses were purified (Discovery Guide Section 6) and amplified (Discovery Guide Section 7) prior to DNA isolation and characterization (Discovery Guide Protocols 8 and 10). They were sequenced at the University of Pittsburgh Bacteriophage Institute with Illumina Sequencing (http://phagesdb.org/phages/) and assembled at the University of Pittsburgh (Newbler and Consed).

The sequences were auto-annotated using DNA Master software. Start sites, reading frames, coding potential, missing or mis-annotated genes, and gene functions were determined using Starterator, Phamerator (Cresawn, et al. 2011), NCBI BLAST (https://www.ncbi.nlm.nih.gov), GeneMark (Besemer et al. 2005), Glimmer (Delcher et al. 2007), Phagesdb (Russell and Hatfull, 2017), and HHPred (Zimmermann et al. 2018, Gabler et al. 2020) directly and as collected in PECAAN (Phage Evidence Collection And Annotation Network). Specific guidelines are outlined in the SEA-PHAGES Bioinformatics Guide (https://seaphagesbioinformatics.helpdocsonline.com/home). We used Aragorn and tRNAscan software (http://mbio-serv2.mbioekol.lu.se/ARAGORN/) to search for tRNA genes. We used Alphafold to analyze DNA polymerase III subunit alpha (DnaE)-like protein (Jumper, et al. 2021, Varadi et al. 2022).

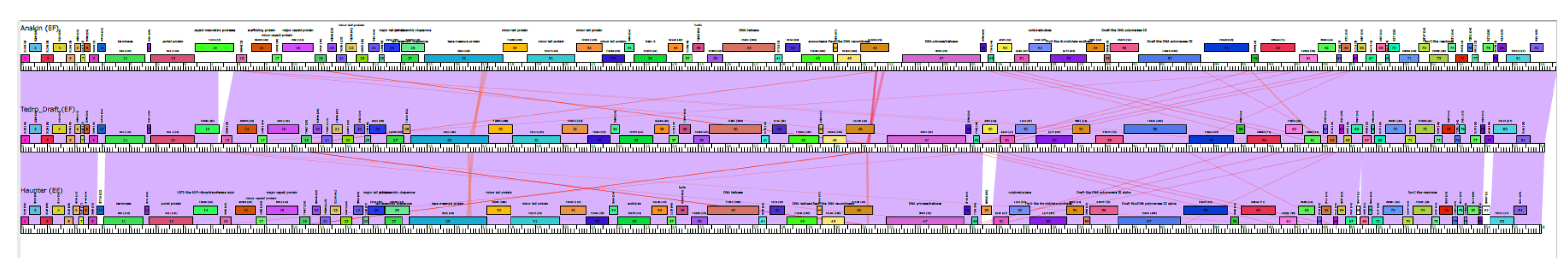


Figure 2. Phamerator map of Tedro and two other Cluster EF phages, Anakin and Haunter, show that Tedro's genome is very similar to other Cluster EF phage genomes.

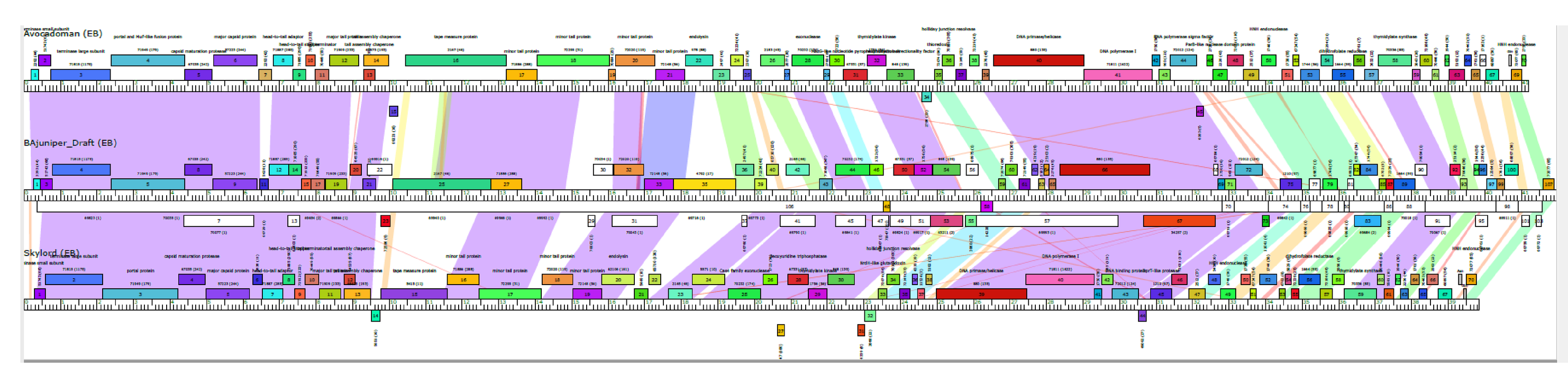


Figure 3. Phamerator map of BAjuniper and two other Cluster EB phages, Avacadoman and Skylord, show that BAjuniper's genome is similar to the genomes of other Cluster EB phages in some regions (purple) but diverges significantly in other regions (white, green, blue, orange). Autoannotation of Bajuniper's genome included many overlapping genes that, upon examination, were not actually genes (white rectangles).

Tedro and most Cluster EF phages have two genes identified by bioinformatics as DnaE-like DNA polymerase III alpha. The two genes differ dramatically in size with the first gene approximately half the size of the second.

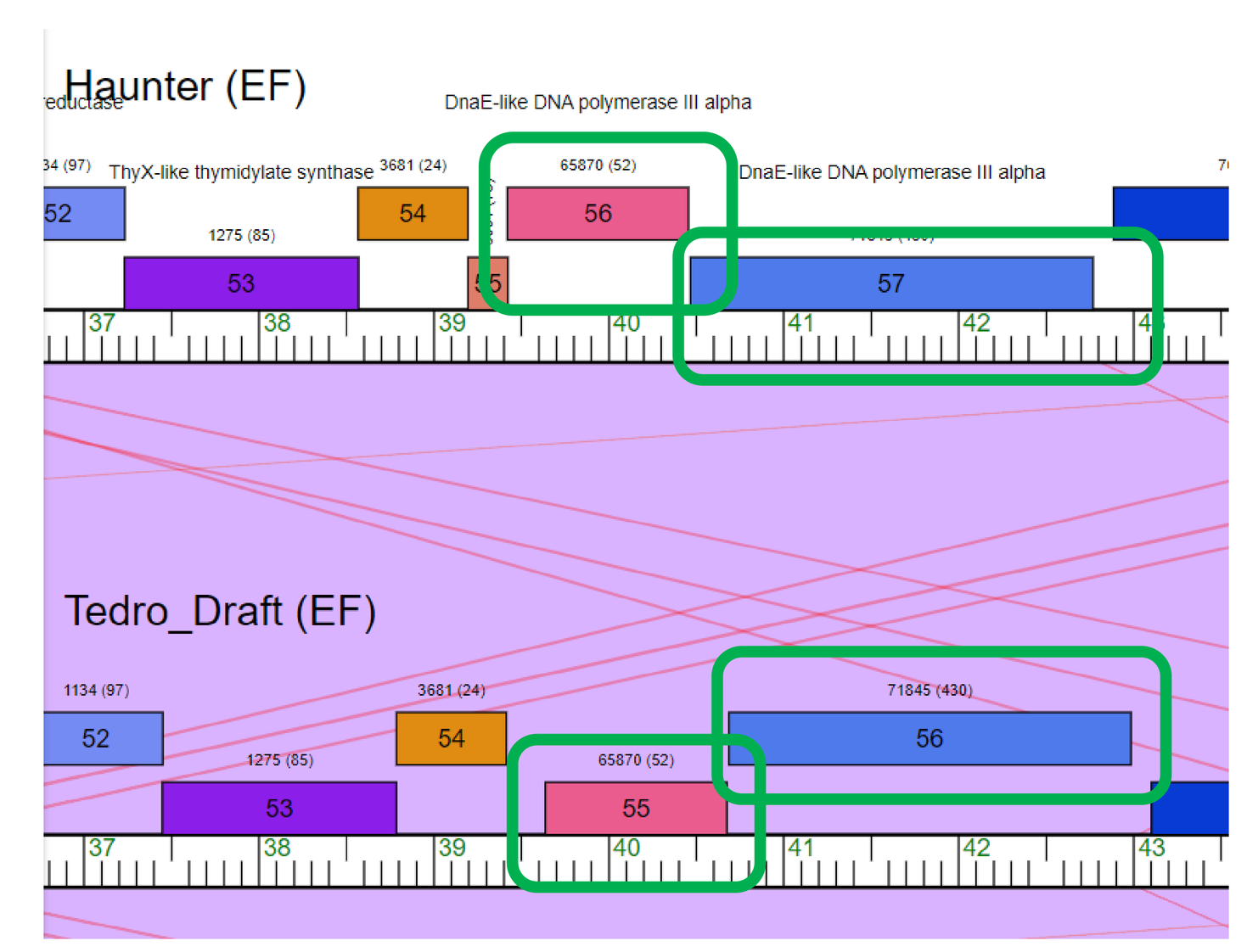
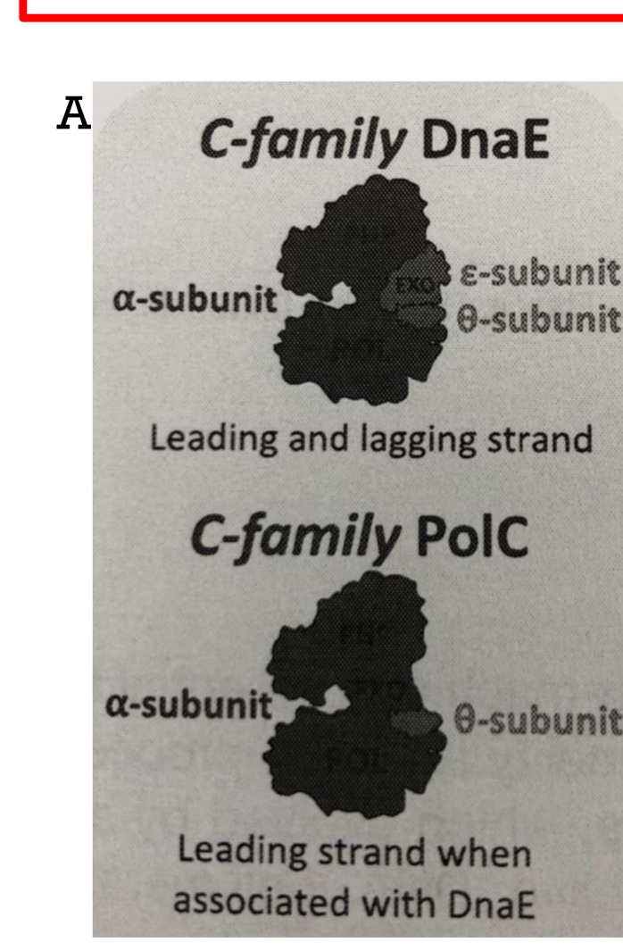


Figure 4. Phamerator map showing the two genes annotated as DnaE-like DNA polymerase III alpha in Haunter and Tedro.

We were curious about these two genes, both called DnaE-like DNA polymerase III alpha, but so different in size. We dug into the literature and found that:

- **DNA polymerase III has three parts**
- **It is responsible for the replication of bacterial genomes**
- **It consists of DNA polymerase (alpha subunit), processivity factor β-clamp, and clamp loader complex**
- **There are two forms of DNA polymerase III (Figure 5a)**
  - **DnaE in *E coli* and others**
  - **PolC in low GC content bacteria (*B. subtilis*)**
- **The polymerase or alpha subunit of both PolC and DnaE has a similar structure including palm, thumb, and finger domains with a PHP domain, oligo binding domain, and a helix-hairpin-helix domain (Figure 5b).**
- ***M. tuberculosis* has two DnaE genes, the second is called DnaE2 and is believed to be involved in error-prone translesion DNA synthesis (Ditse, et al. 2017)**



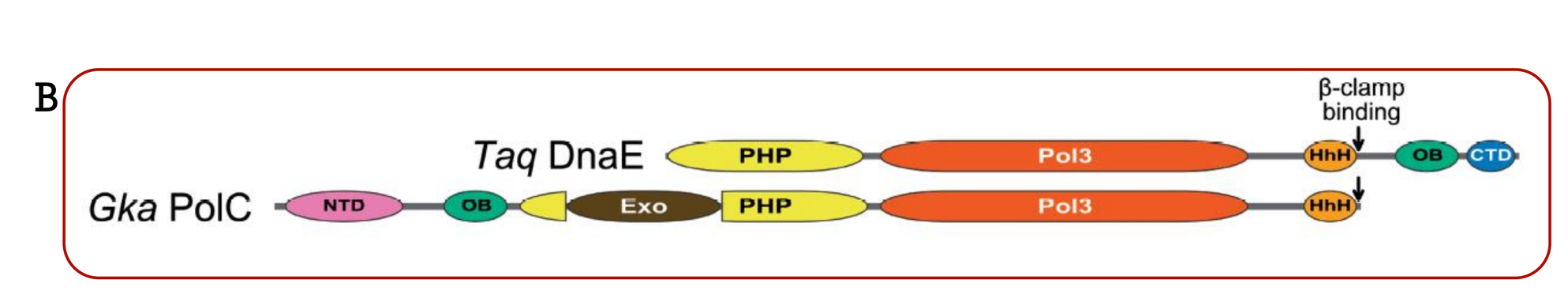


Figure 5. Two types of DNA polymerases: DnaE and PolC. A. 3D models of DnaE and PolC (Raia et al. 2019). B. Linear organization of DnaE and PolC domains (Timinskas et al. 2013).

We compared structures for Tedro_56 and Tedro_57 to structures for DNA III alpha subunit from various organisms by running sequences through the AI-driven structure prediction software, Alphafold (Junper, et al. 2021, Varadi et al. 2022) and looked carefully at the HHPred analysis (Zimmermann et al. 2018, Gabler et al. 2020).

- **Tedro_56's predicted 3D structure is similar to the crystal structure of the first half of DNA polymerase from *M. tuberculosis* and even more similar to the predicted structure for DnaE-like DNA polymerase III alpha from *Clostridium perfringens*.**
- **Tedro_57's predicted 3D structure is similar to the crystal structure of DNA polymerase from *M. tuberculosis*.**
- **HHPred analysis of Tedro_56 suggests Tedro_56 has domains that resemble PHP and polymerase and histidinol phosphatase domains with 100% probability and 93% coverage.**
- **HHPred analysis of Tedro_57 suggests that Tedro_57 has domains that resemble nucleotide transferase, PHP, NTPase, alpha subunit finger, helix-hairpin-helix, and oligonucleotide binding domains with 100% probability and more than 98% coverage.**

We postulate that Tedro_57 may be a DnaE-like DNA polymerase III alpha. Tedro_57 is 31% identical to DnaE-like DNA polymerase III alpha subunit from *M. tuberculosis* with 50% similarity. Tedro_56 may be a truncated version of this gene that could function in DNA repair, like DnaE2, with structural similarity to DnaE from *C. perfringens.* These two proteins are 30% identical and show 47% amino acid similarity.

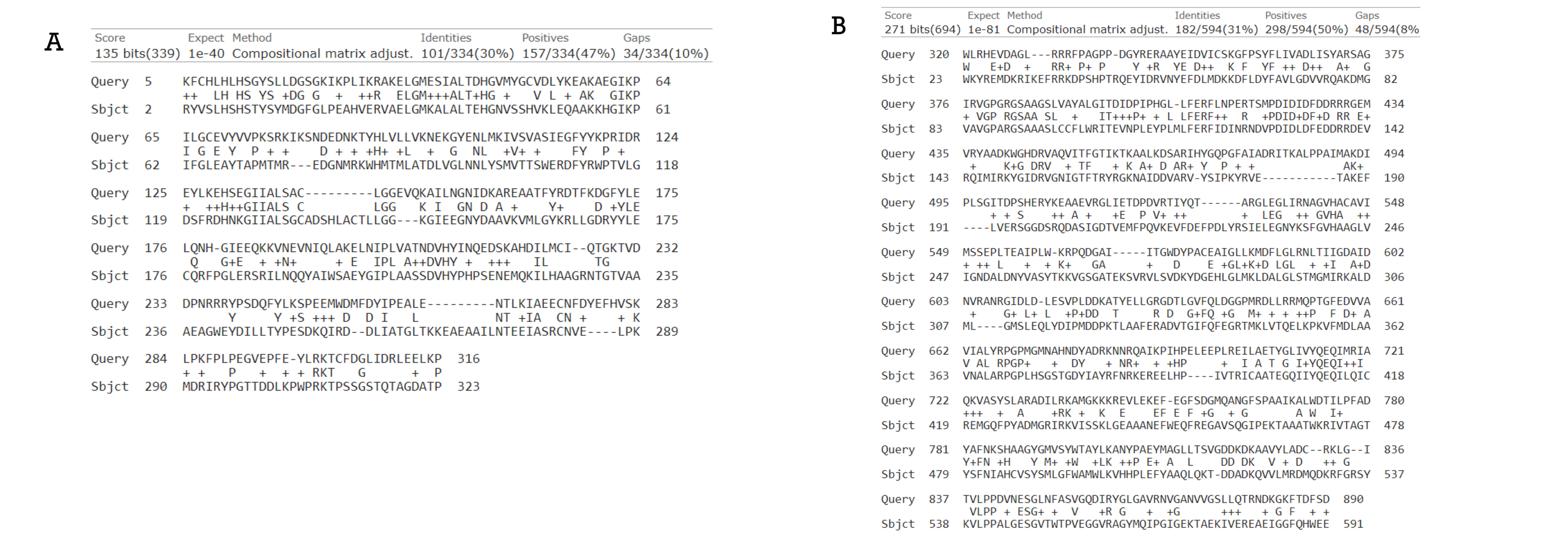


Figure 6. NCBI BLAST amino acid alignment. A. The amino acid sequence for Tedro_56 aligned to DnaE from *C. perfringens* shows 30% identify and 47% amino acid similarity. B. Tedro_57 aligned to DNA polymerase from *M. tuberculosis* shows 31% identify to DnaE-like DNA polymerase III alpha subunit from *M. tuberculosis* and 50% similarity.

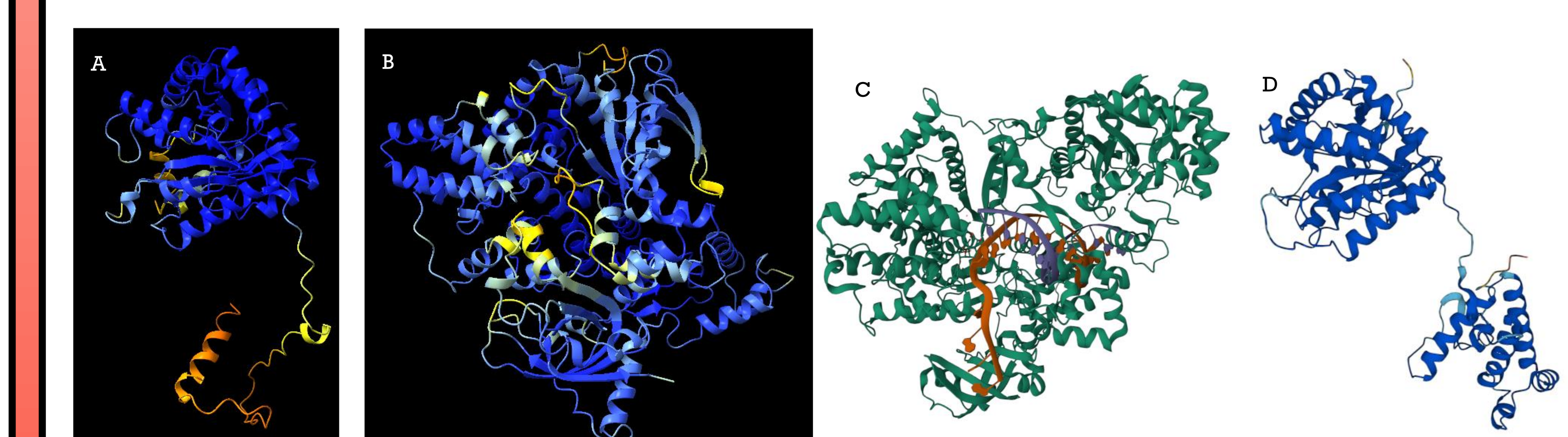


Figure 7. Experimental and predicted 3D structures. A. Alphafold predicted 3D structure of Tedro_56. B. Alphafold predicated 3D structure of Tedro_57. C. TEM structure of *M. tuberculosis* DnaE1. D. Alphafold predicted 3D structure of DNA polymerase III, subunit alpha (DnaE)-like protein from *Clostridium perfringens*.

Discussion and Future Directions

We are finalizing our annotation of BAjuniper and are preparing the files to submit to GenBank for publication. We have already submitted the files for our annotation of Tedro to GenBank and are anxious to see our work published.

We are working on writing a *Microbiology Resource Announcement*—another peer-reviewed publication reporting the discovery, characterization, and annotation of both viruses. We expect that paper will be published sometime this summer or early fall.

We will also be presenting our work at three scientific meetings: the Northwestern College Celebration of Research, the Iowa Academy of Science Annual Meeting, and the SEA-PHAGES Symposium, all in April 2023.

References

Besemer J, Borodovsky M. Nucleic Acids Res 33:W451–W454 (2005).
Cresawn SG, et al. BMC Bioinformatics 12:395 (2011).
Delcher AL, et al. Bioinformatics 23:673–679 (2007).
Ditse et al. Microbiol Spectr. 2017 March ; 5(2) (2017).
Gabler F, et al. Curr Protoc Bioinformatics. 72(1):e108 (2020).
Jumper, J et al. Nature (2021).
Kasman, L. M., & Porter, L. D. https://www.ncbi.nlm.nih.gov/books/NBK493185/ (2022).
Keen E. C. BioEssays, 37(1), 6–9 (2015).
Mirdita, M., et al. Nat Methods 19, 679–682 (2022).
Russell DA, Hatfull GF. Bioinformatics 33:784–786 (2017).
Schneider, C. A et al. Nature Methods, 9(7), 671–675 (2012).
Strange, J.E.S., et al. Sci Rep 11, 1600 (2021).
Varadi, M et al. Nucleic Acids Research (2022).
Zimmermann L, et al. Mol Biol. (2018).

Howard Hughes Medical Institute