



The University of Victoria
December 5th, 2025

University
of Victoria

CLASS:
CSC427
Bioinformatics and Clinical
Applications

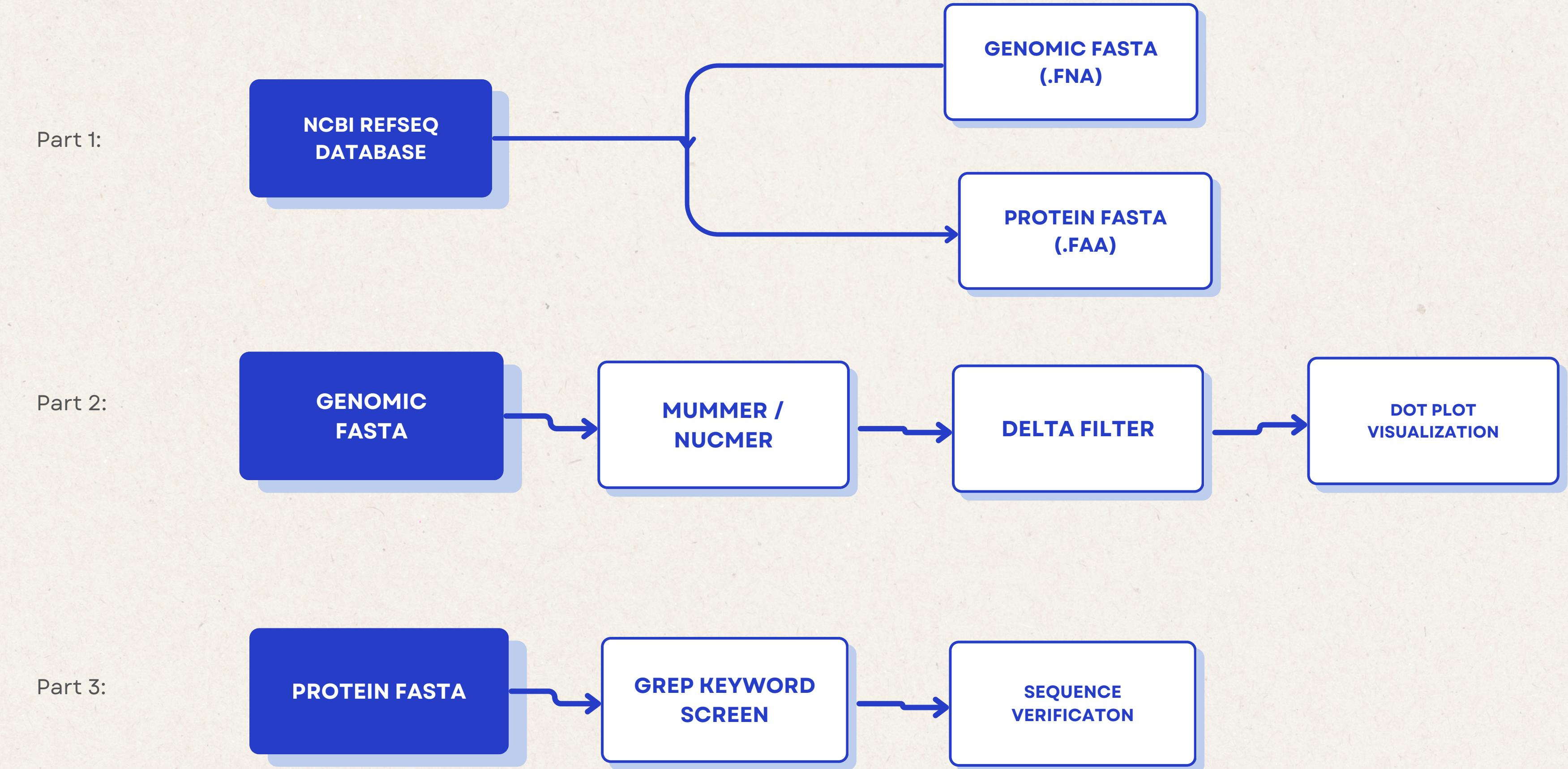
PRESENTED BY:
Kacey Friesen
V00978721

PRESENTED TO:
Ibrahim Numanagic + others

GENOMIC PLASTICITY IN LEPTOSPIRA INTERROGANS

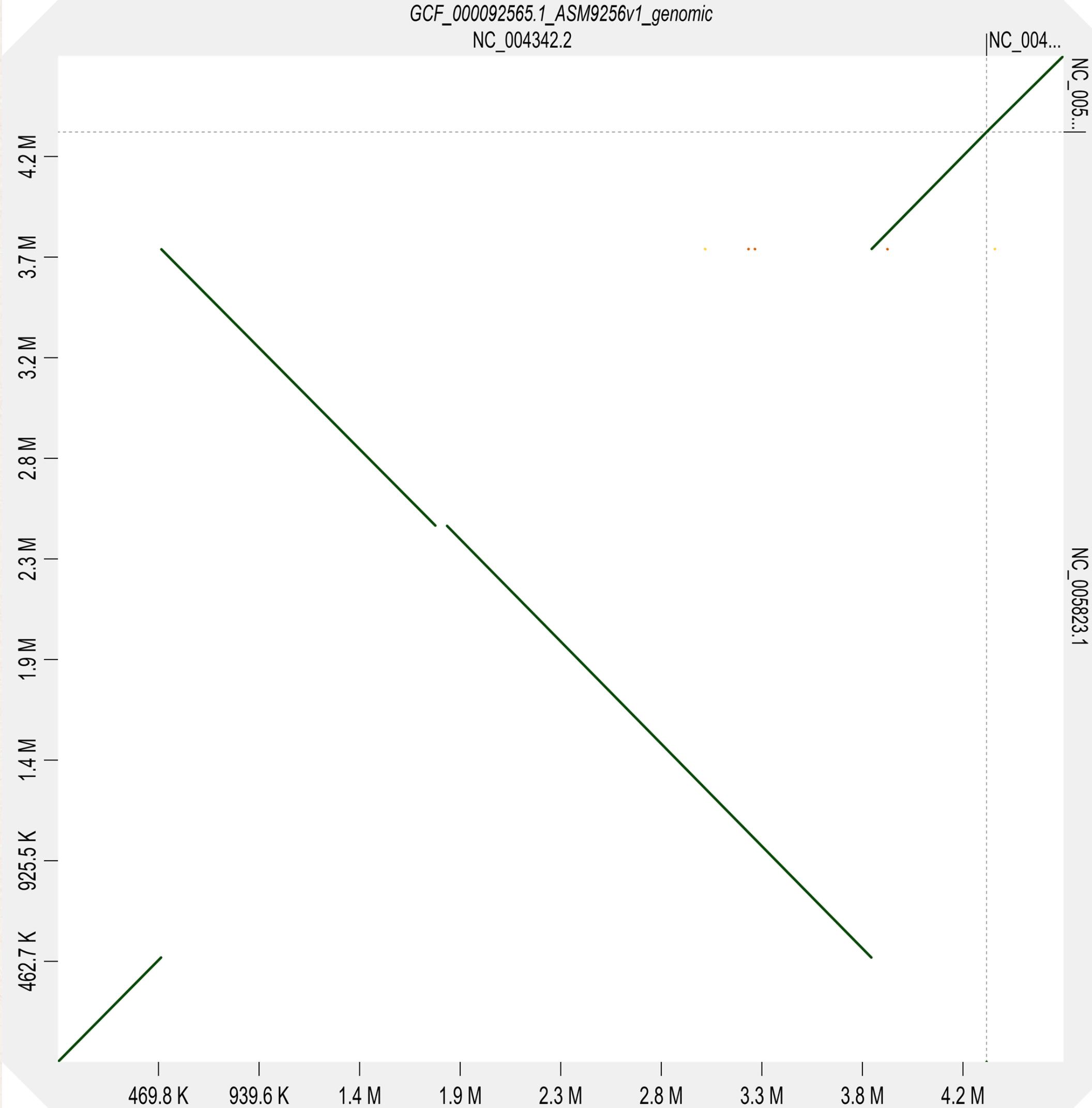
Leptospira interrogans serovars Copenhageni (rats) and Lai (mice) exhibit distinct host specificities. In 2004, Nascimento et al. proposed that a massive chromosomal inversion and the loss of the LigA adhesin in Lai drive these differences. We computationally replicated this study using 2025 RefSeq assemblies to validate these structural and functional claims.

Pipeline



Structural Results

Figure 1: Whole-genome structural alignment generated via D-GENIES. The plot compares the genomic architecture of serovar Copenhageni (X-axis) against Lai (Y-axis). The central diagonal break and negative slope indicate a large-scale inversion event, flanked by repetitive IS1501 insertion sequences which likely mediated the rearrangement.



Re-Evaluating Lig Proteins

Target Gene	2004 Paper Claim	2025 Replication Result
LigB	Present	Present
LigA	Absent	Present

Background

The Pathogen: *Leptospira interrogans* causes Leptospirosis, a severe zoonotic disease affecting kidneys/liver.

The Mystery: Why do genetically similar serovars (>95% identity) infect different hosts?

The 2004 Study: Comparing the first complete genomes of Copenhageni and Lai revealed:

1. A large inversion in Chromosome 1
2. A “broken” virulence profile in Lai - Missing LigA

Algorithm 1: Whole Genome Alignment

Core Technology

Core Technology: We utilized NUCmer (NUCleotide MUMmer), which employs a Suffix Tree data structure for rapid, large-scale genome alignment.

1. Seeding: The algorithm identifies Maximal Unique Matches (MUMs)—exact subsequences shared between both genomes that are unique in both.
2. Clustering: MUMs are clustered to form alignment anchors.
3. Extension: Gaps between anchors are filled using a modified Smith-Waterman dynamic programming algorithm.

Why This Tool?

Unlike local aligners (BLAST), NUCmer is optimized for global alignment of closely related bacterial species (>95% identity), making it ideal for detecting large structural rearrangements like inversions.

Structural Analysis

Analysis for Figure 1

Whole-genome alignment of Lai (Y-axis) vs. Copenhageni (X-axis)

Synteny

The diagonal line indicates high sequence identity (>95%).

Inversion

The distinct negative slope (green line) between coordinates ~1.5 Mb and 3.0 Mb confirms a massive chromosomal inversion.

Mechanism

This region is flanked by IS1501 elements, acting as recombination breakpoints.

Sequence Evidence

- To prove the Lai LigA result was not a database error, we extracted the sequence.
- The sequence contains a valid Start Codon (M) and is full-length, confirming it is a functional protein, not a pseudogene.

```
>WP_002093809.1 lipoprotein adhesin LigA [Lai]
MKKIFCISIFLSMFFQSCMSWPLLTSLAGLAAGK...
```

Hypothesis

Hypothesis

The hypothesis is that modern bioinformatics tools will confirm the structural instability (inversions) but may correct historic proteomic errors due to improved sequencing quality

Objectives

Replicate Structure - Validate the Chromosome I inversion using whole-genome alignment (MUMmer).

Re-assess Virulence - Screen modern proteomes to verify the reported absence of LigA in Serovar Lai.

Algorithm 2: Handling Repetitive Elements

The Challenge

Leptospira genomes are plagued by Insertion Sequences (IS1501)—highly repetitive mobile genetic elements. Standard alignment often breaks or filters these out, obscuring the inversion boundaries.

The Solution (--maxmatch)

- We executed NUCmer with the --maxmatch flag.
- Technical Detail: This forces the suffix tree traversal to report all matches, regardless of uniqueness. This ensures that the repetitive IS elements flanking the inversion are included in the alignment map, rather than being discarded as repeats.

Validation (D-GENIES)

- To rule out assembly artifacts (scaffold misordering), we cross-validated with D-GENIES.
- Mechanism: Uses minimap2 to map query contigs to the reference, then computationally reorders the query scaffolds to maximize diagonal synteny, reducing visual "noise."

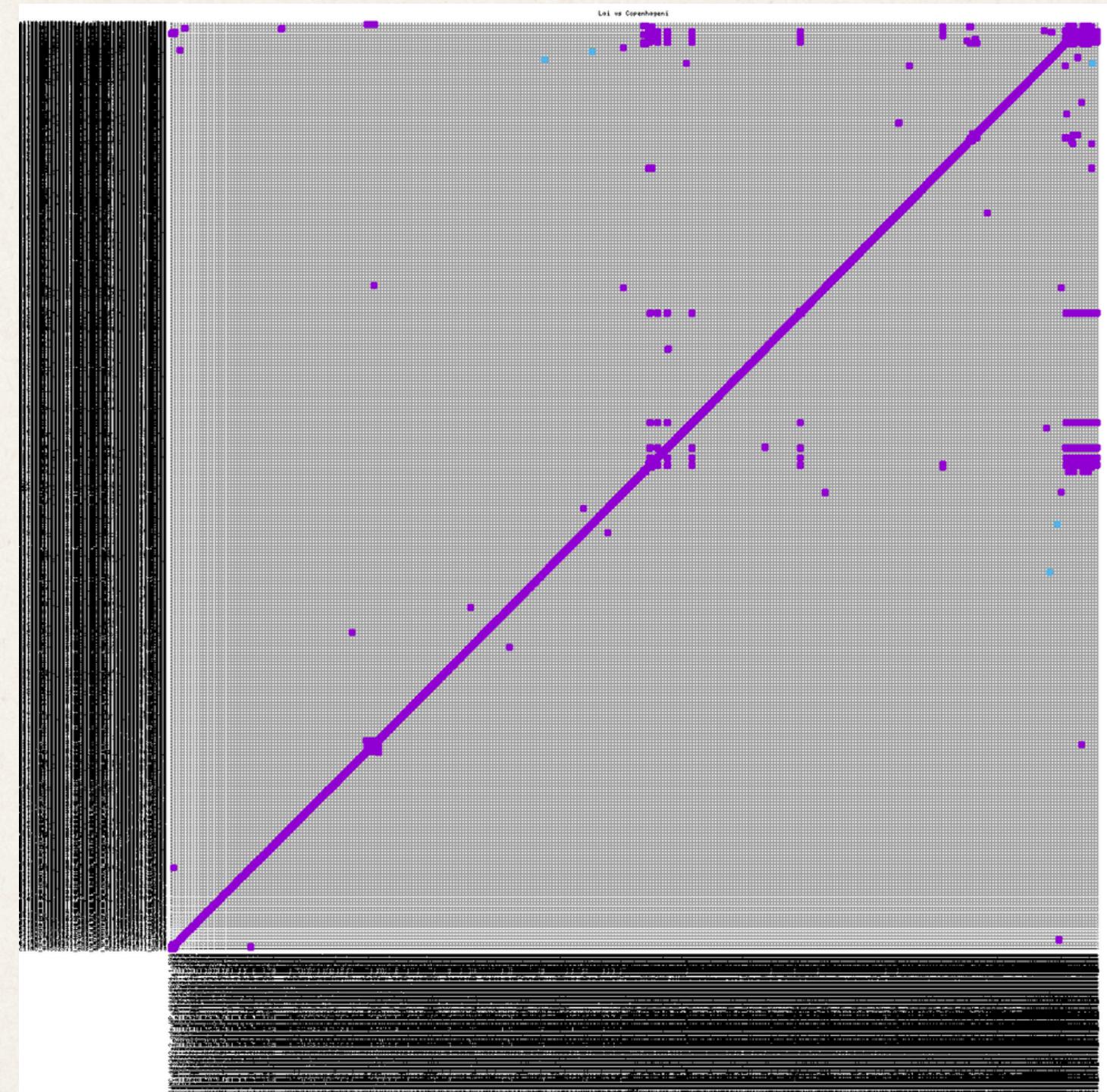


Figure 2.

Raw whole-genome alignment generated via CLI mummerplot

Algorithm Details & Raw Data

Initial Output - Figure 2

The raw NUCmer output (purple plot) reveals the complexity of using draft RefSeq assemblies. The "scattered" noise along the axes represents unsorted genomic scaffolds.

Refinement

While the raw plot successfully detected matches, the lack of syntenic ordering obscured the large-scale inversion.

Scaffold Sorting

To resolve this, D-GENIES (seen in Figure 1) was used, which algorithmically reorders query scaffolds to minimize diagonal breaks, revealing the true 1.5 Mb inversion event clearly.

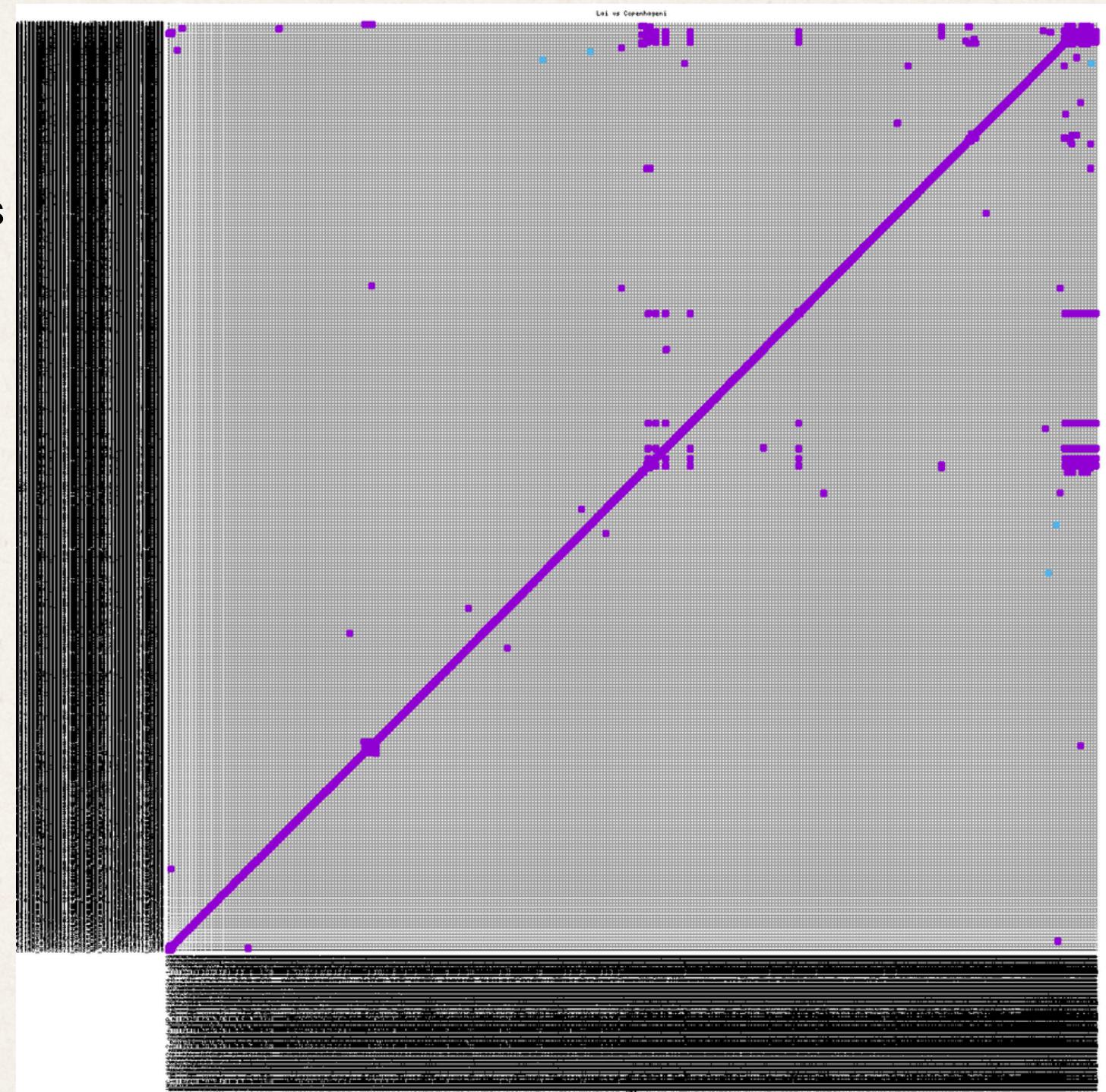


Figure 2.

Raw whole-genome alignment generated via CLI mummerplot

Proteome Mining Strategy

METHODOLOGY

- Matching: We employed Regular Expression (Regex) filtering on the annotated proteome (.faa) to identify specific adhesin motifs.
- Target: grep -i "immunoglobulin-like" and grep "Lig[AB]"

NOISE FILTERING

Raw hits were manually curated to exclude DNA Ligase (EC 6.5.1.2), a DNA repair enzyme often falsely flagged by simple "Lig" keyword searches.

SEQUENCE INTEGRITY CHECK

Extracted the N-terminal region of the Lai LigA hit to verify the presence of a Methionine (M) start codon and open reading frame (ORF) continuity, ruling out pseudogene fragmentation.

Conclusion

SUMMARY:

1. STRUCTURE REPLICATED: THE CHROMOSOMAL INVERSION IS A ROBUST FEATURE OF SEROVAR LAO

2. PROTEOME UPDATED: SEROVAR LAI POSSESSES A FUNCTIONAL LIGA, CONTRADICTING THE 2004 STUDY

SIGNIFICANCE:

HISTORIC GENOMIC CONCLUSIONS MUST BE RE-EVALUATED AS ASSEMBLY QUALITY IMPROVES. LAI IS MORE LIKELY TO BE VIRULENT THAN PREVIOUSLY ASSUMED

1. Nascimento, A. L., et al. (2004). Comparative genomics of two *Leptospira interrogans* serovars reveals novel insights into physiology and pathogenesis. *Journal of Bacteriology*, 186(7), 2164-2172. <https://doi.org/10.1128/jb.186.7.2164-2172.2004>
2. Marçais, G., et al. (2018). MUMmer4: A fast and versatile genome alignment system. *PLoS computational biology*, 14(1), e1005944. <https://doi.org/10.1371/journal.pcbi.1005944>