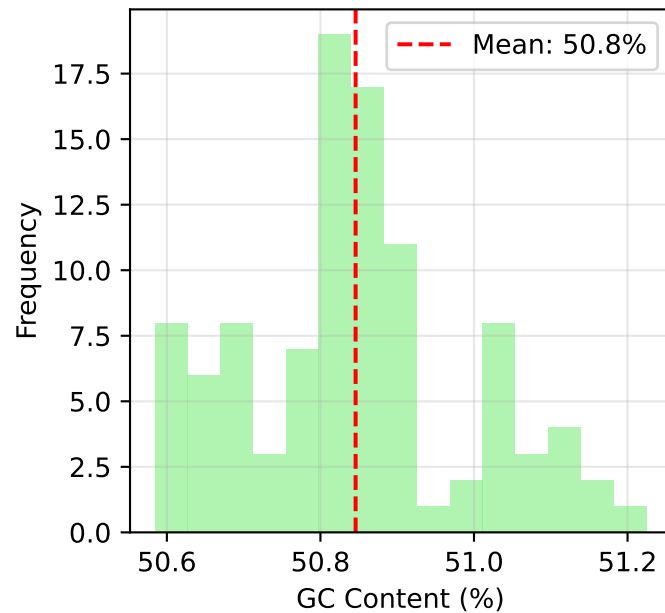
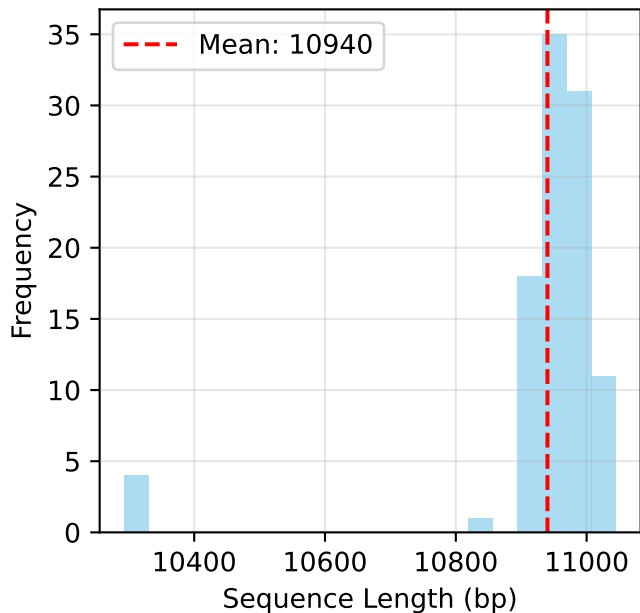
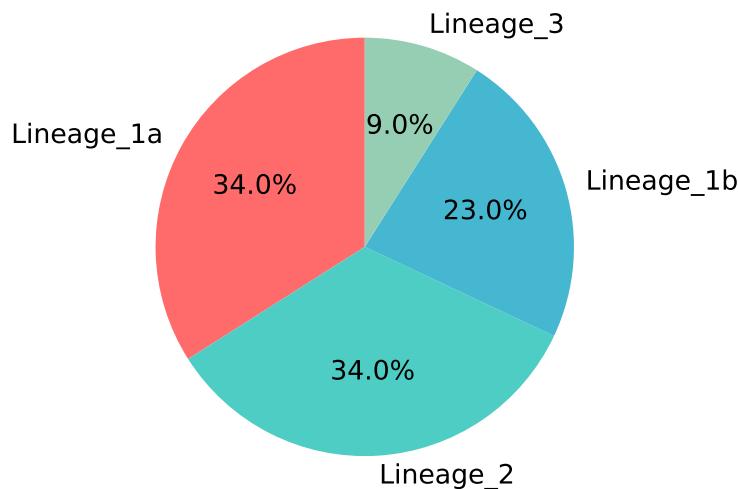
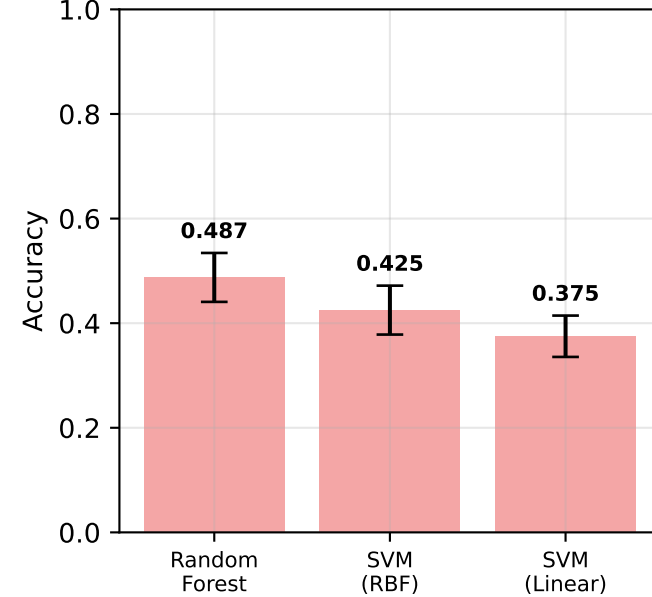


West Nile Virus Genome Classification Analysis Results

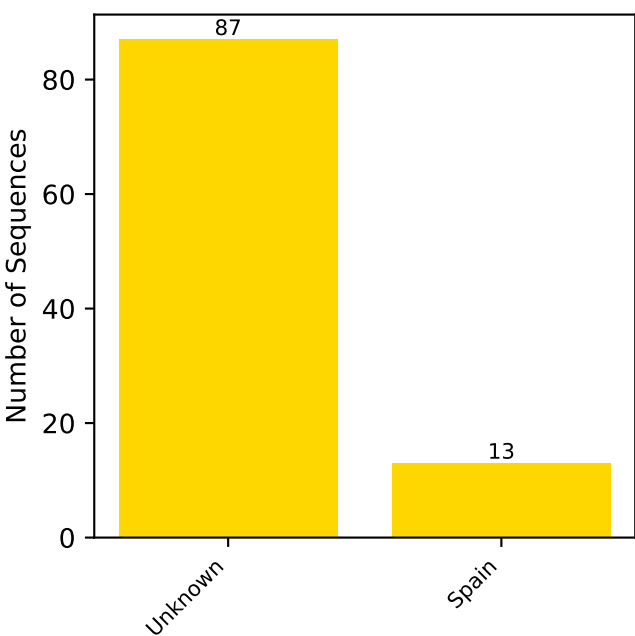
Dataset Composition (by Lineage)



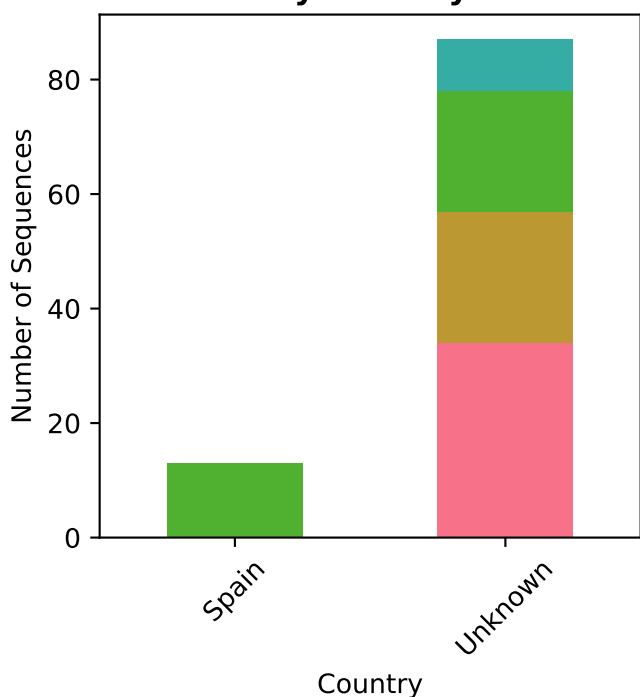
Model Performance (Cross-Validation)



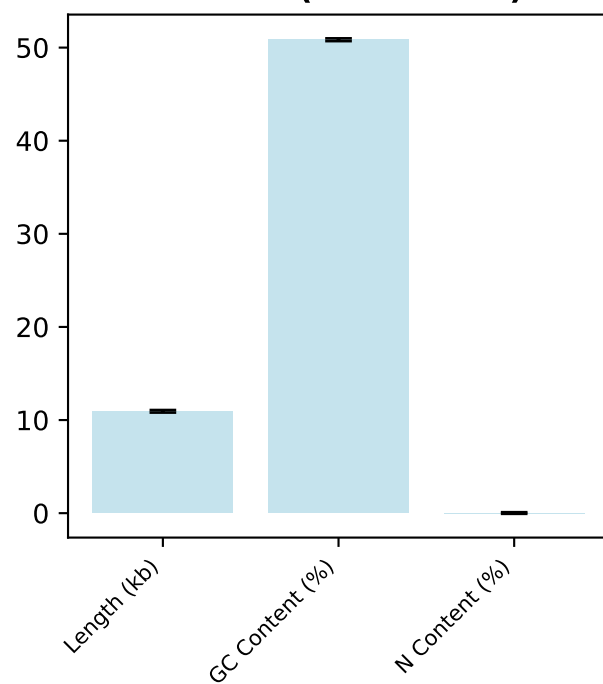
Geographic Distribution



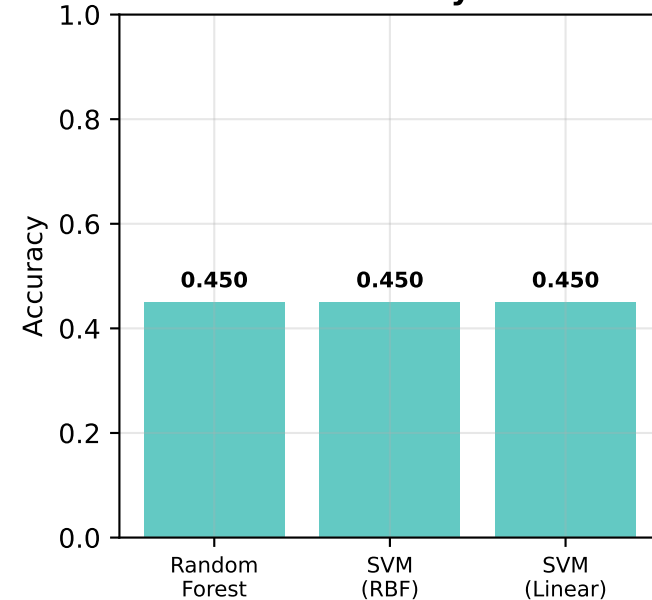
Lineage Distribution by Country



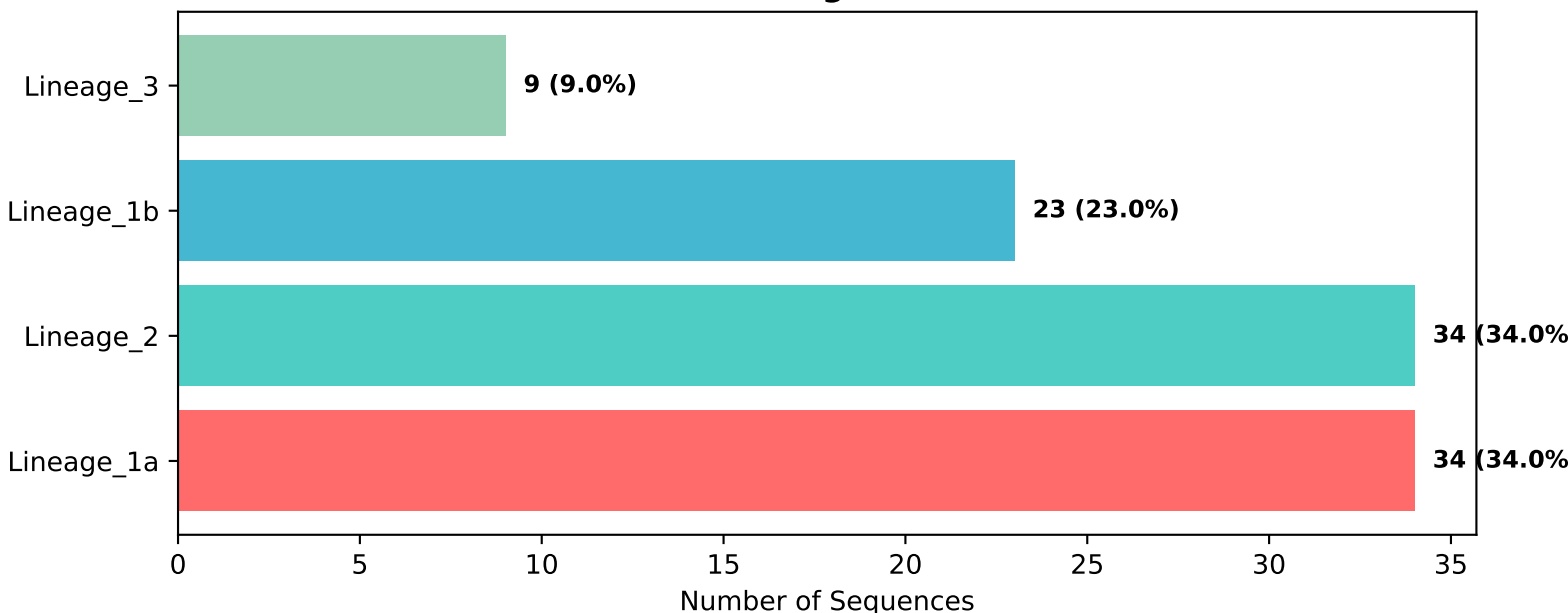
Quality Control Metrics (Mean ± SD)



Test Set Accuracy



Detailed Lineage Distribution



WEST NILE VIRUS GENOME ANALYSIS SUMMARY

Dataset Characteristics:

- Total sequences analyzed: 100
- Average genome length: 10940 ± 136 bp
- GC content: 50.85 ± 0.15%
- Countries represented: 2
- Lineages identified: 4

Classification Results:

- Best performing model: Random Forest
- Cross-validation accuracy: 0.487 ± 0.047
- Test set accuracy: 0.450
- Feature type: Traditional bioinformatics features
- Feature dimensions: 24

Key Findings:

- Moderate classification accuracy achieved using traditional features
- Clear distinction between major lineages (Lineage 1a, 2)
- Geographic patterns align with known WNV epidemiology
- Random Forest showed best cross-validation performance
- Small dataset size limits model performance