



AfricaCDC
Centres for Disease Control
and Prevention



KLEBGEN PROJECT DATA ANALYSIS REPORT

Moise Christian Jr. Meka

August 19th, 2025



Agenda

- A. Completed activities
- B. Samples collection
- C. Analysis workflow
- D. TheiaProk results
- E. KPSC analysis results
- F. *E. coli/Shigella* analysis results



COMPLETED ACTIVITIES

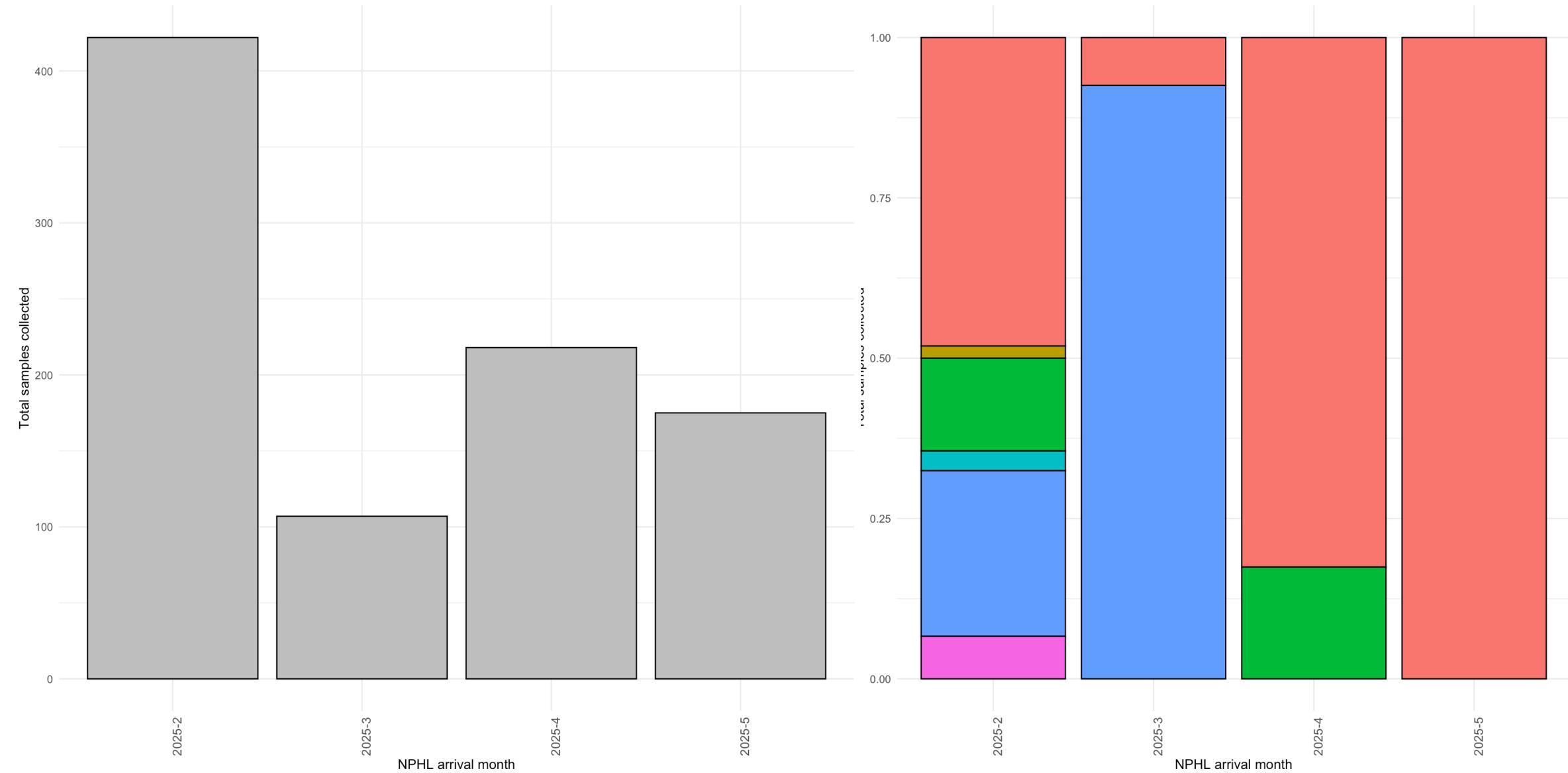
- **Isolates and metadata collection from health facilities**
 - 922 strains : 322 *E. coli*, 498 *Klebsiella* and 102 *Staphylococcus*
 - Sex ratio : 1.09 F/M
 - 82 (8.89%) neonates' strains : 26 *E. coli*, 43 *Klebsiella* and 13 *Staphylococcus*
 - Blood samples : 19.87%
- **Microbiology species verification**
 - 427 strains have been confirmed at the Microbiology lab: 180 are *E. coli*, and 247 are *Klebsiella*.
- **Sequencing**
 - 379 strains were sequenced: 187 of which are *Klebsiella*, 162 are *E. coli*, and 30 are other species or quality control (QC) failed samples.
- **Sharing sequencing results and metadata in Terra**
 - 293 sequences : 127 *E. coli* and 166 *Klebsiella*



SAMPLES COLLECTION

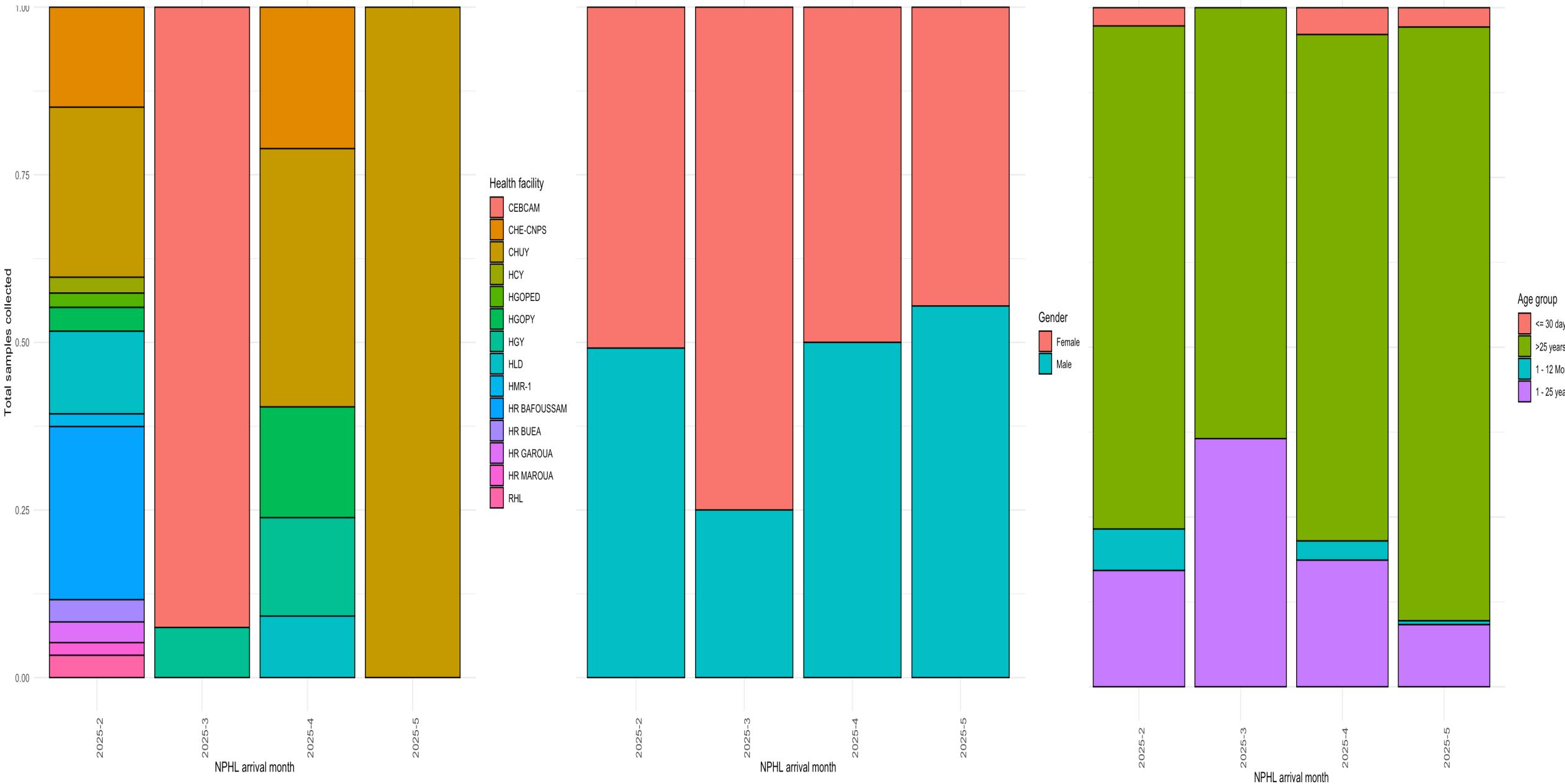


NPHL strains collection dynamics

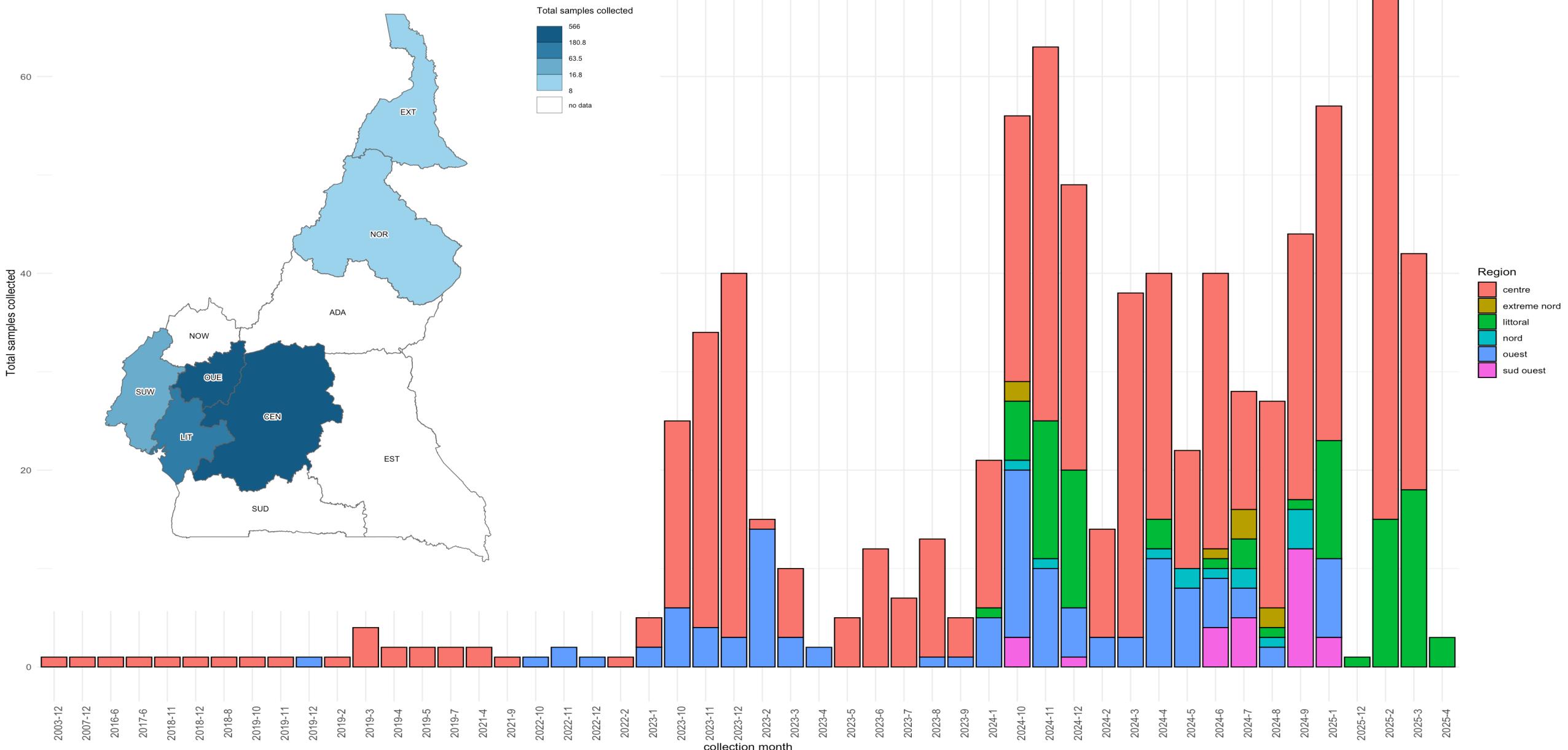




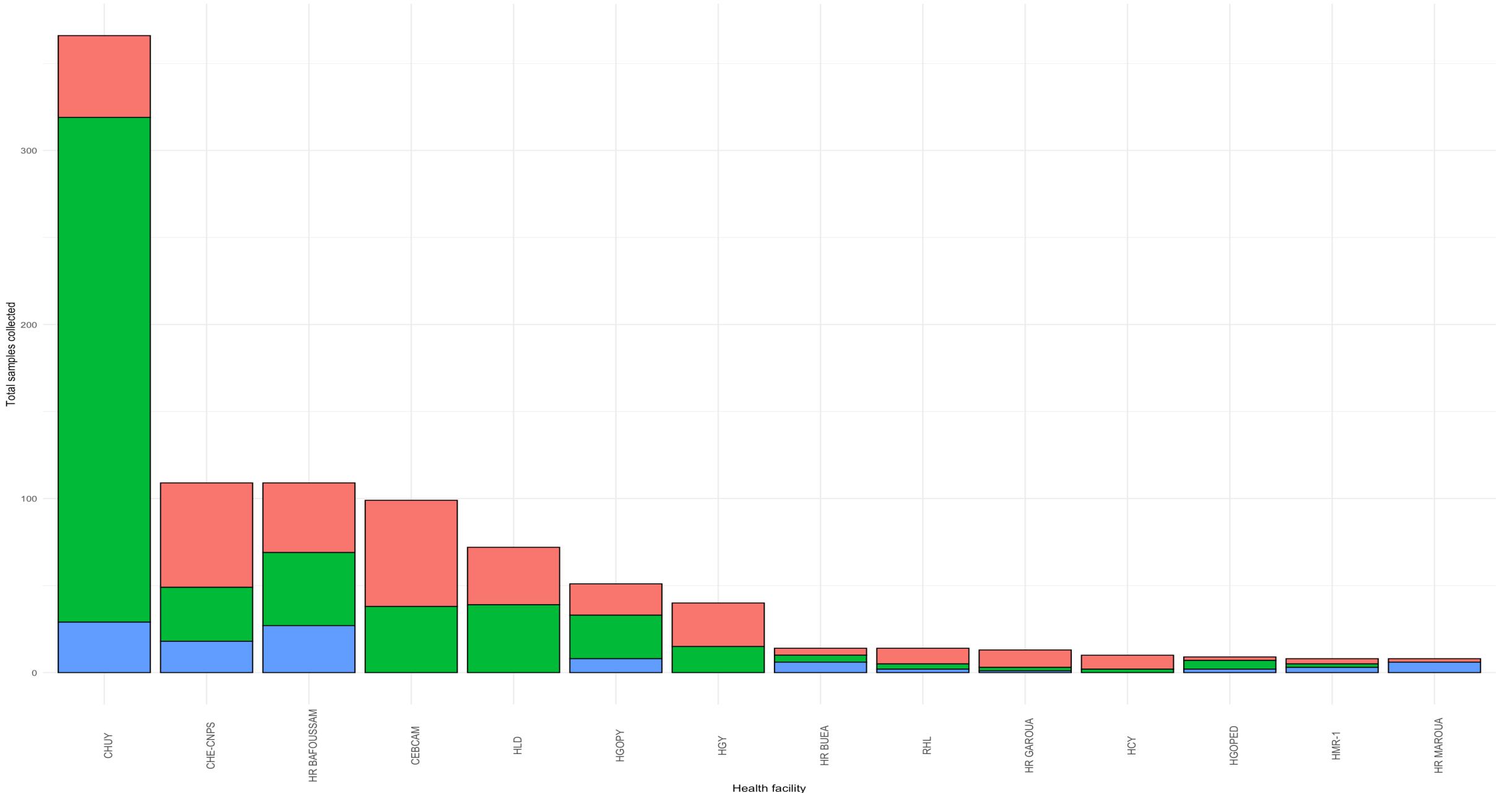
NPHL strains collected details



Regional sample collection dynamics

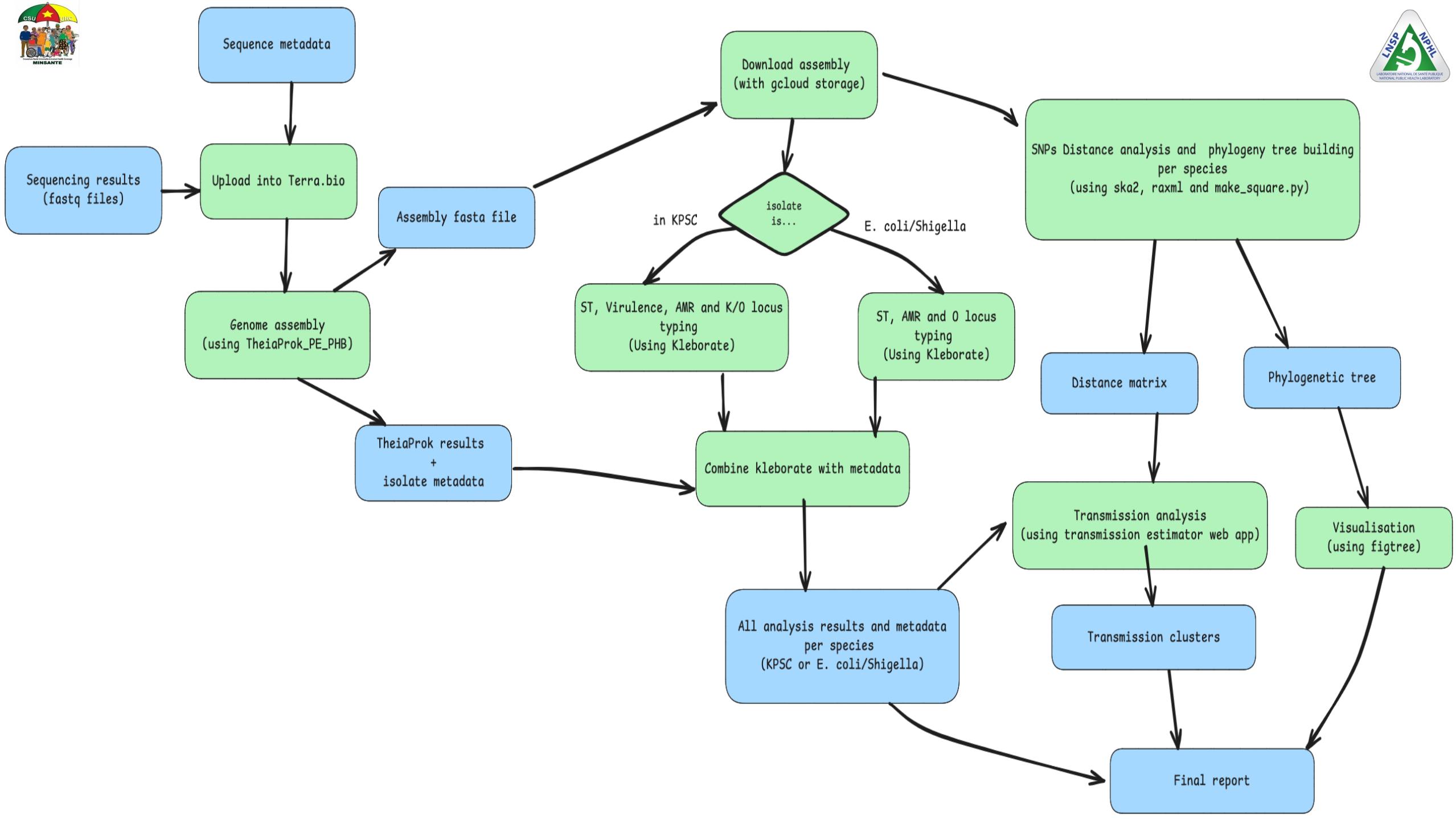


Isolates distribution per Health facility and strain





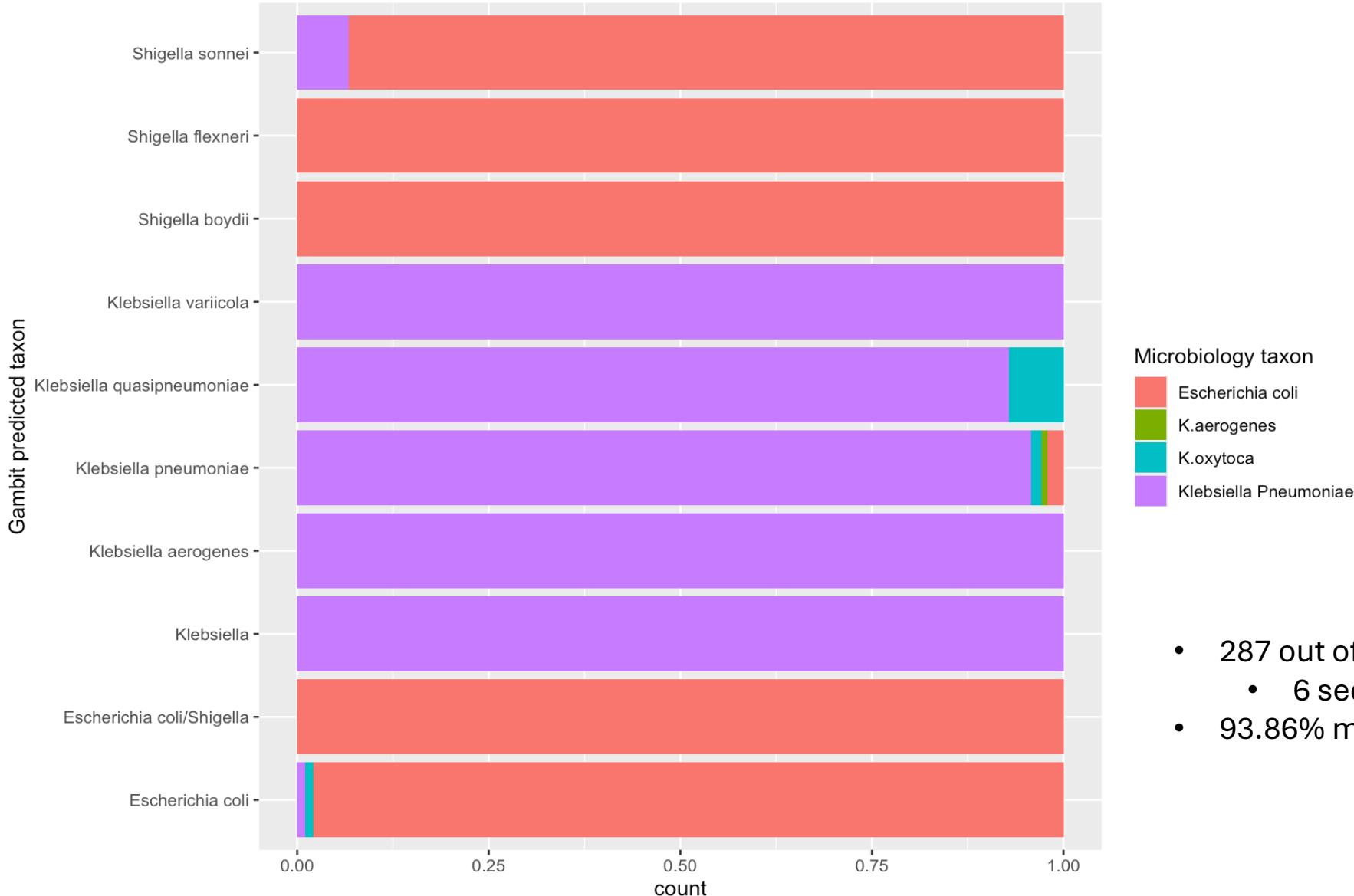
ANALYSIS WORKFLOW





THEIAPROK RESULTS

Comparison of TheiaProk and Microbiology Results

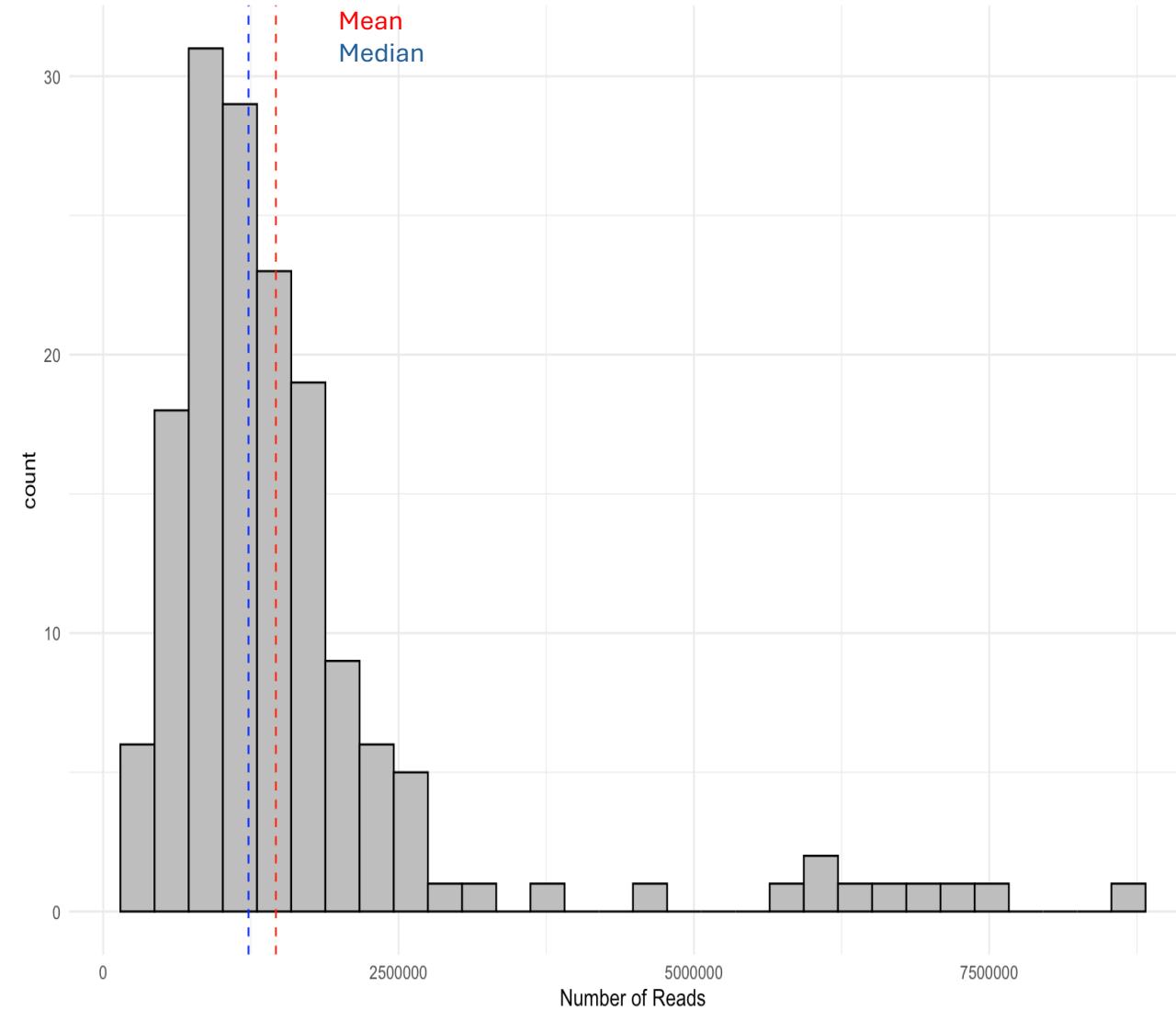


Microbiology taxon

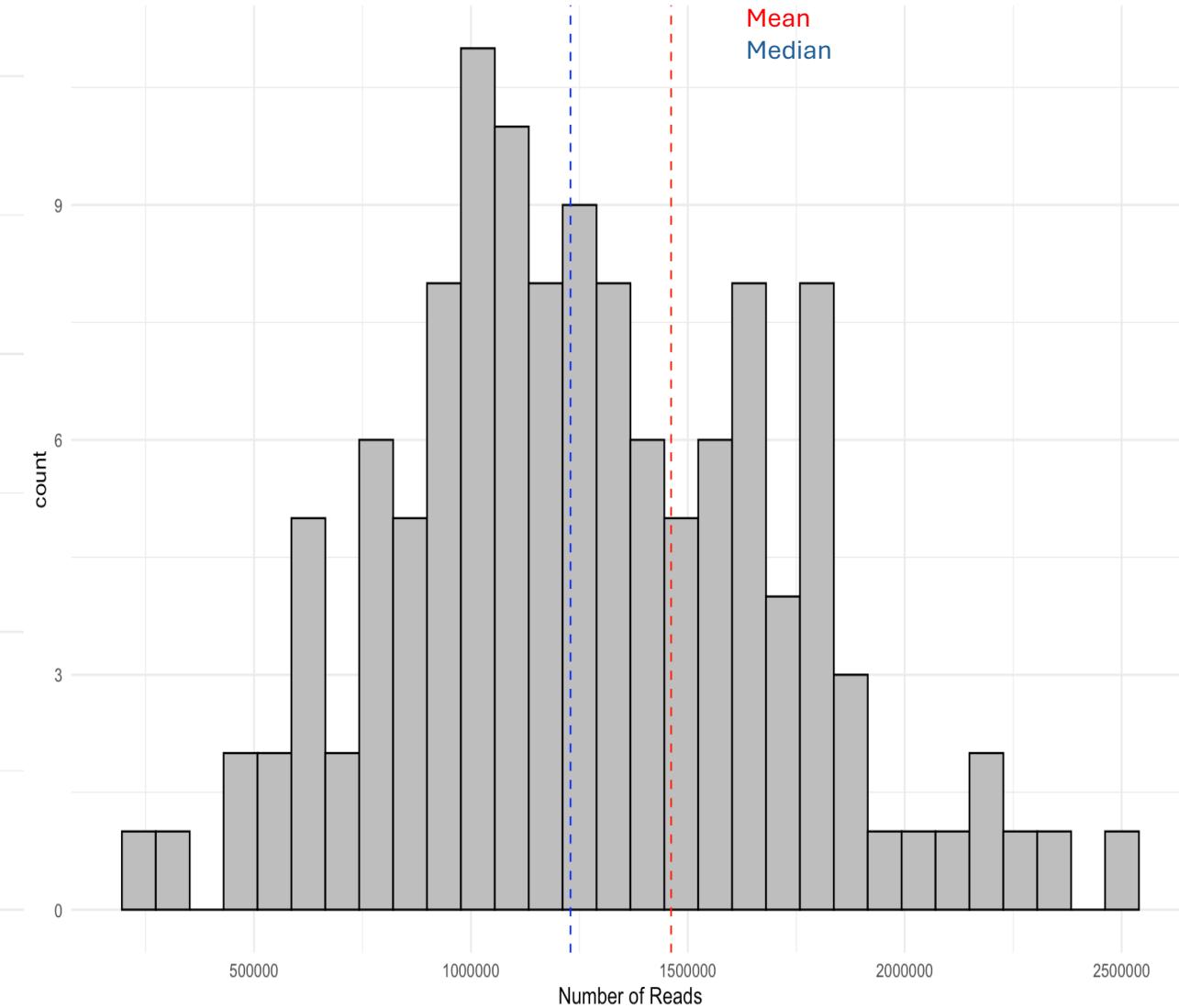
- Escherichia coli
- K.aerogenes
- K.oxytoca
- Klebsiella Pneumoniae

- 287 out of 293 successfully analyzed
 - 6 sequences failed TheiaProk QC
- 93.86% match with microbiology results

Raw sequence quality (Number of reads)

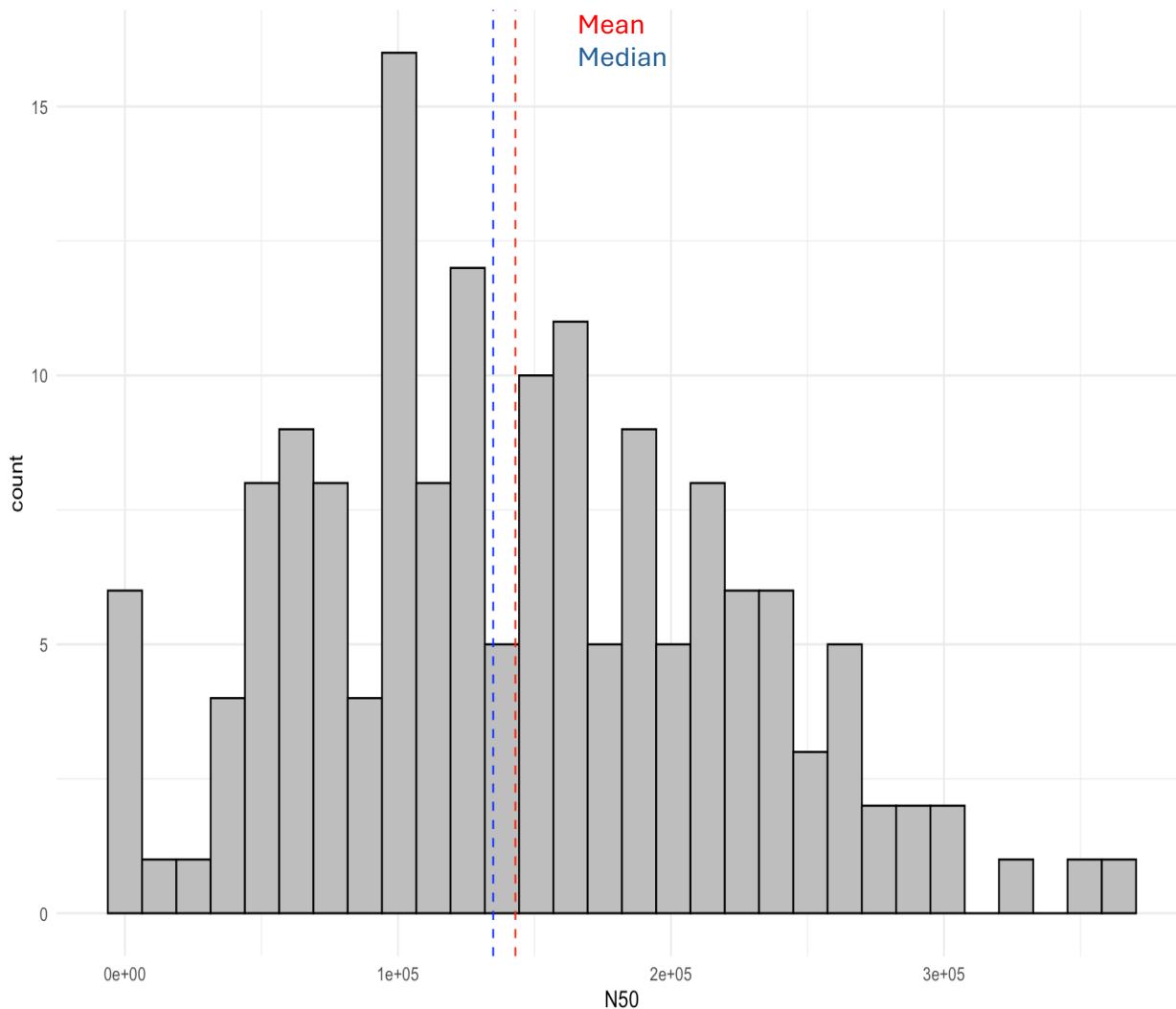


a) KPSC Sequences

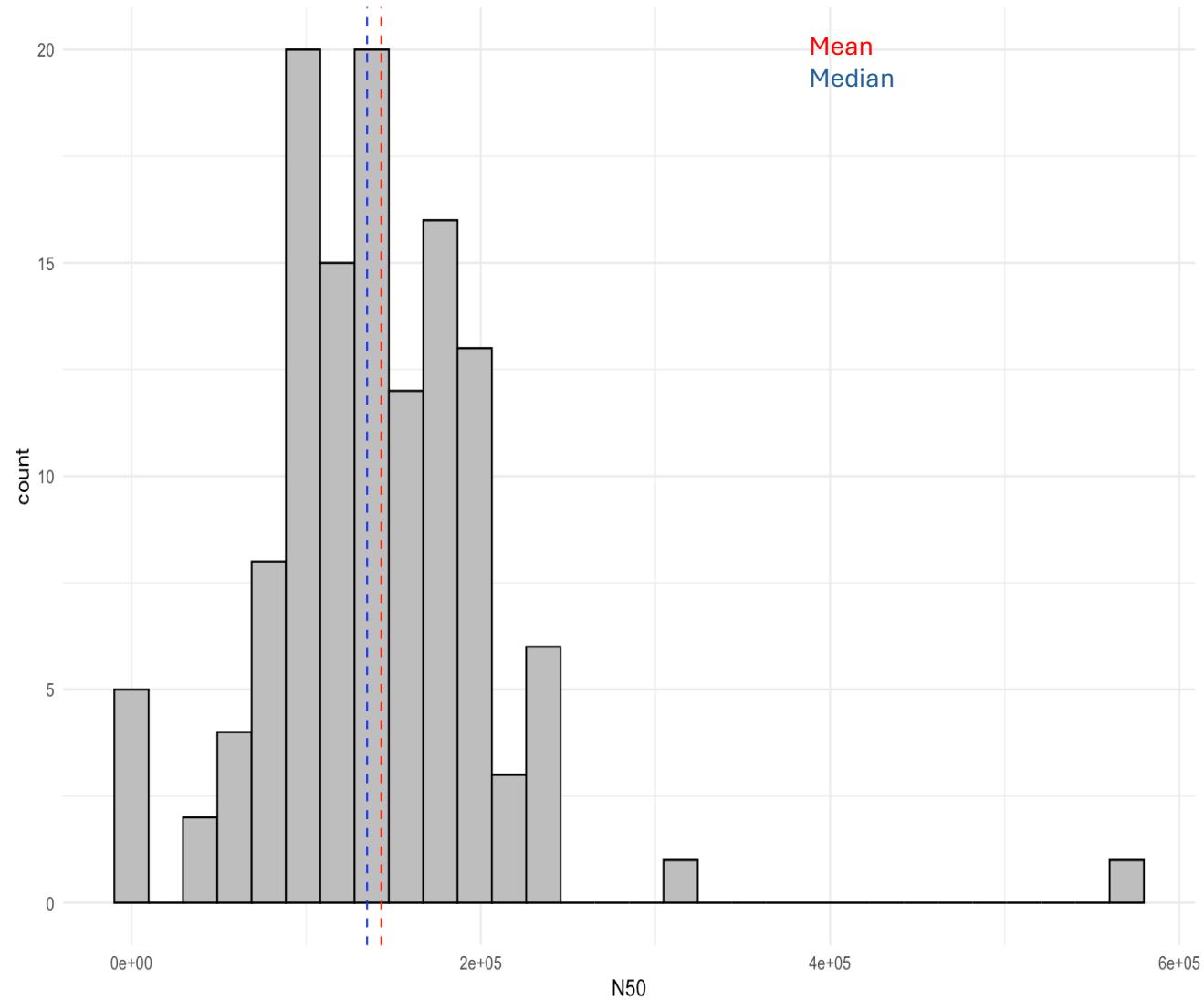


b) *E. coli*/*Shigella* Sequences

Sequence assembly quality (N50)



a) KPSC Sequences

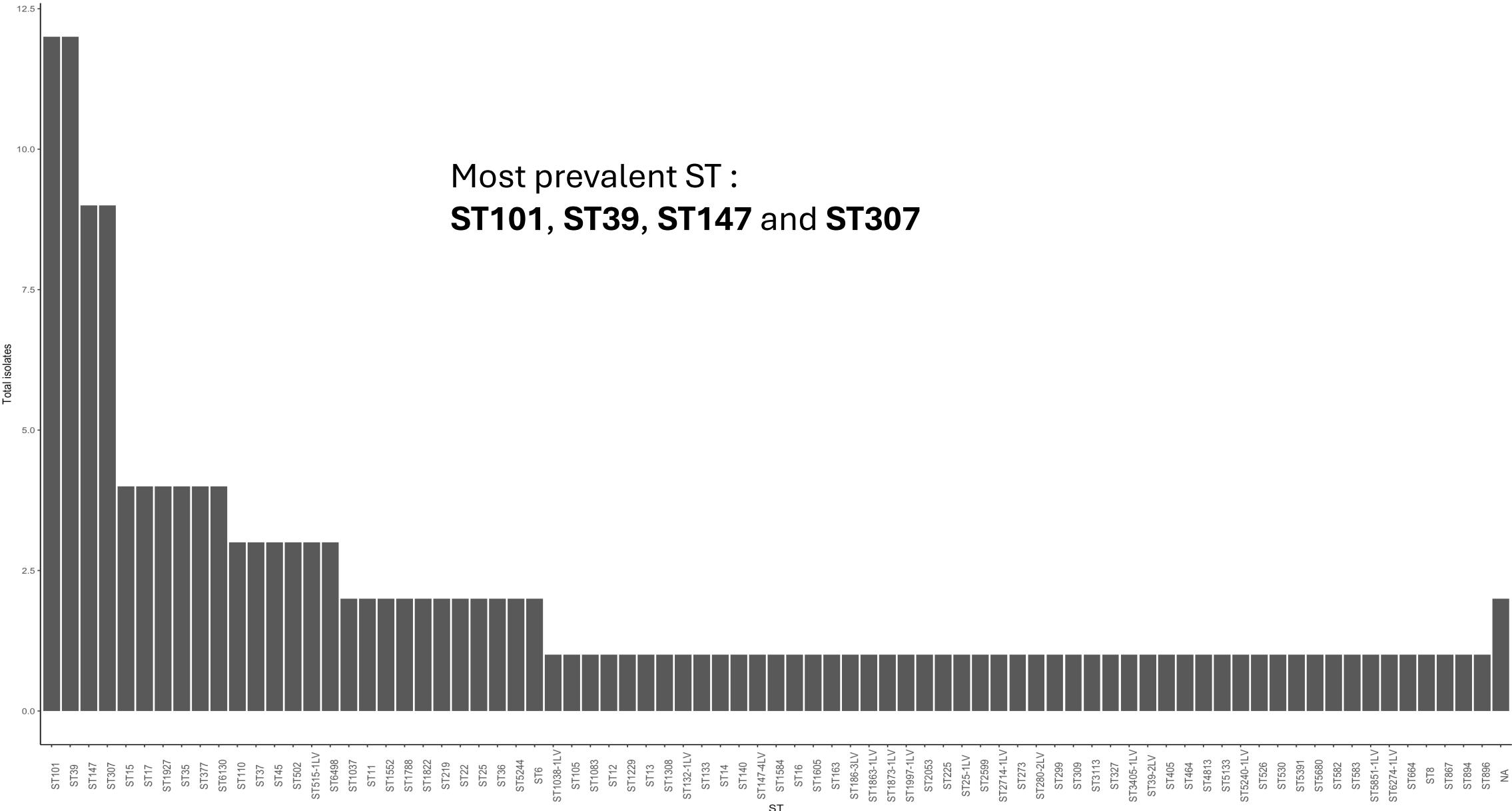


b) *E. coli/Shigella* Sequences

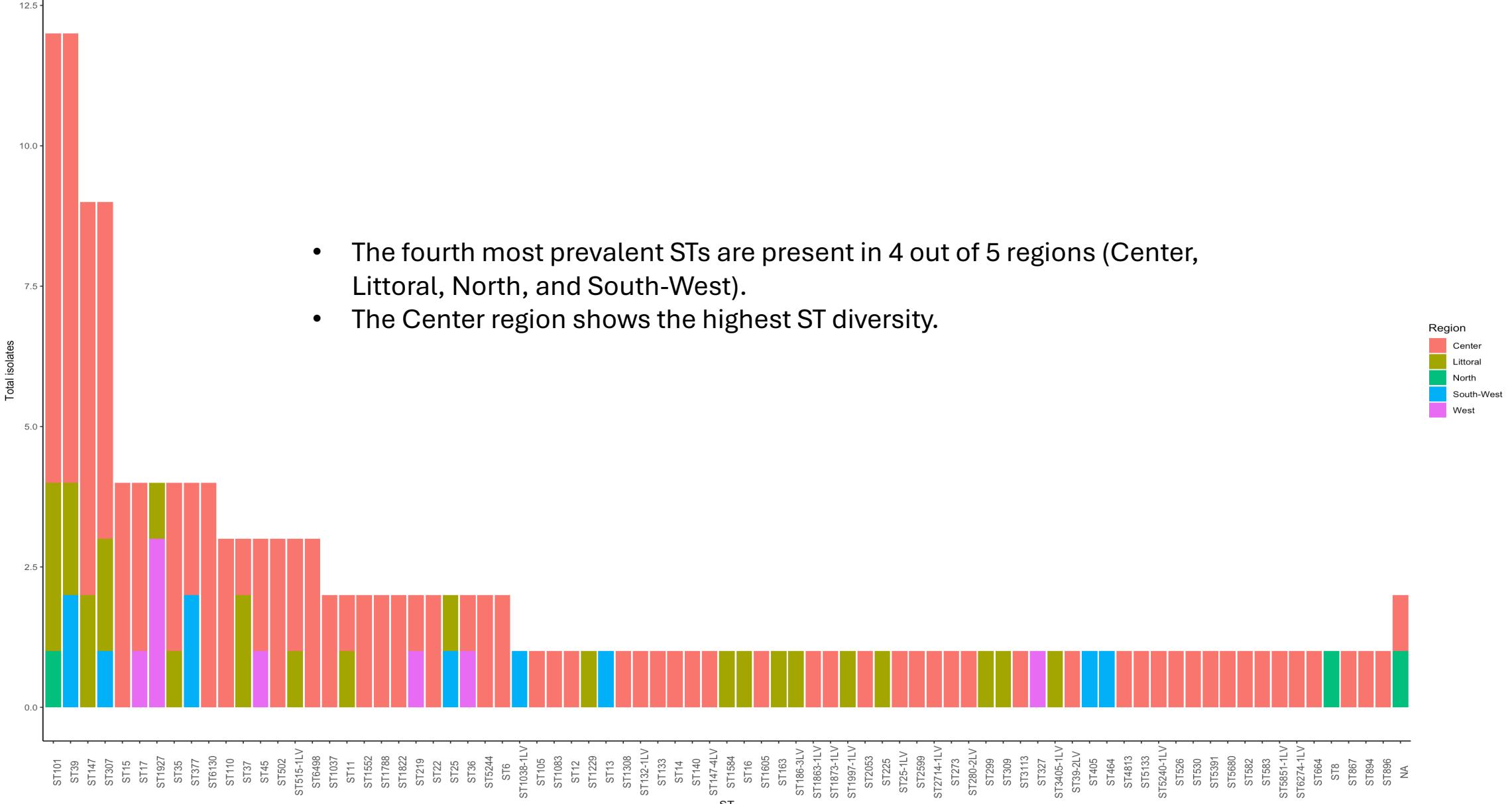


KPSC ANALYSIS RESULTS

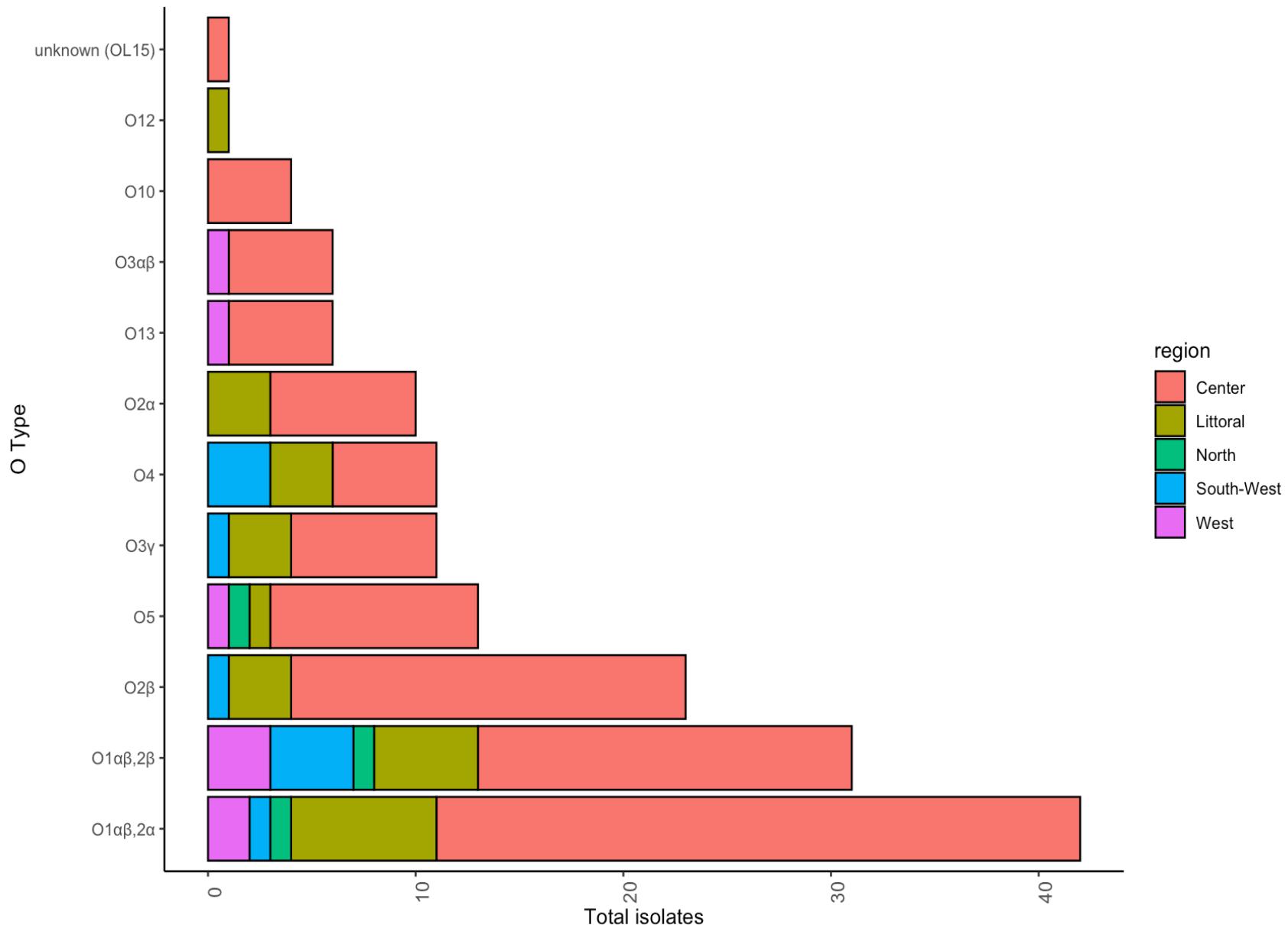
ST prevalence distribution



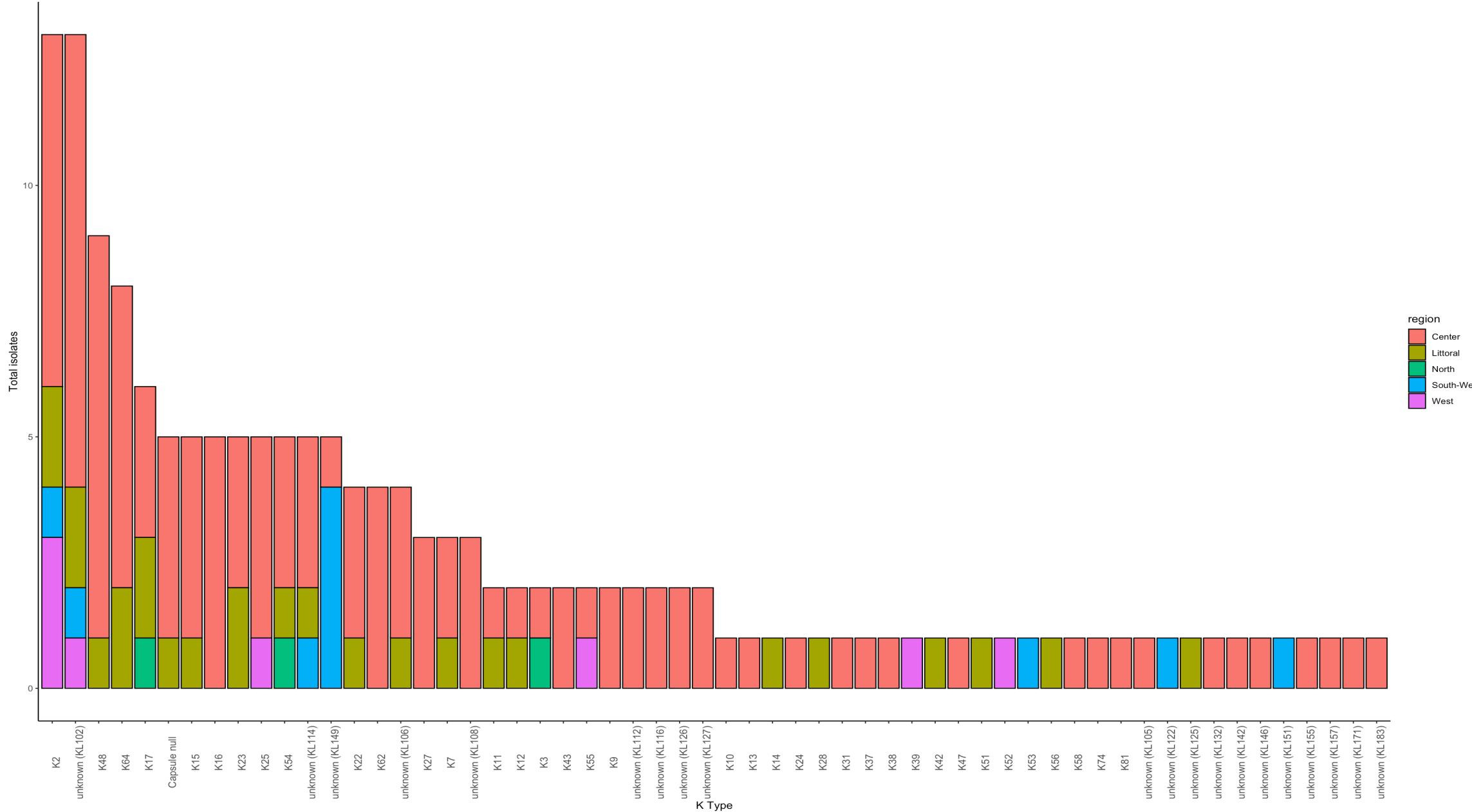
ST prevalence distribution per region



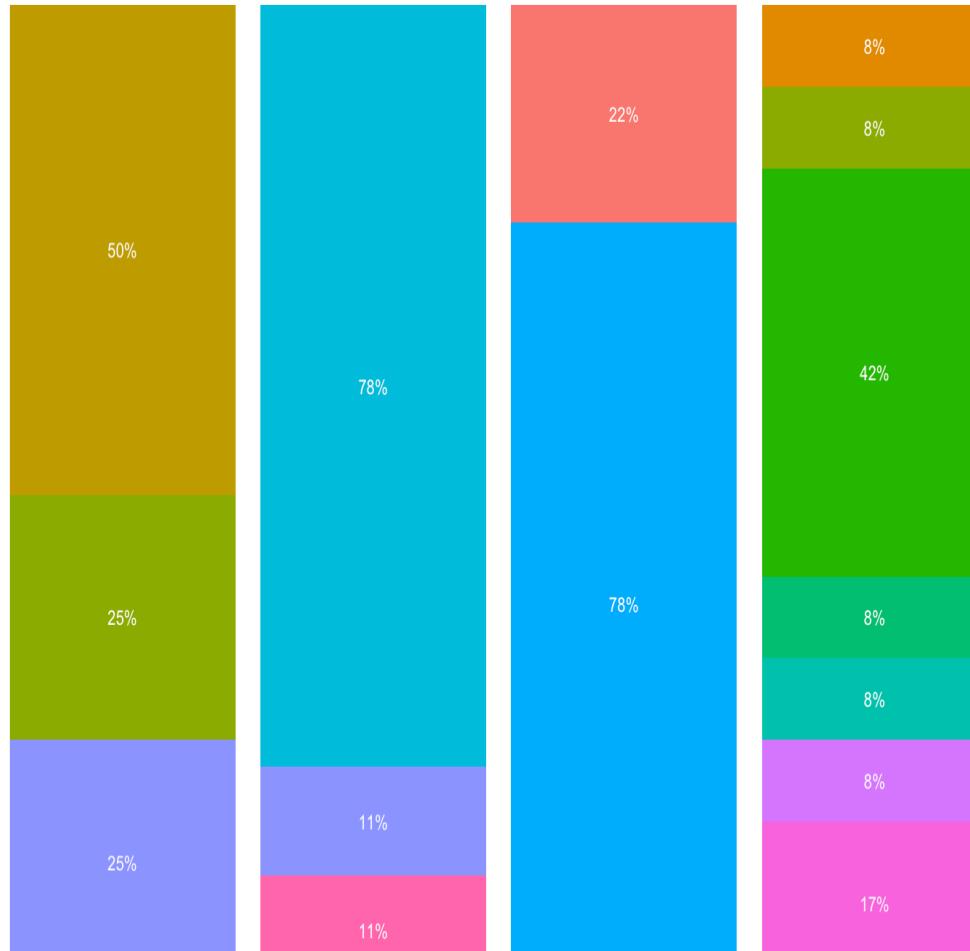
O type Distribution per region



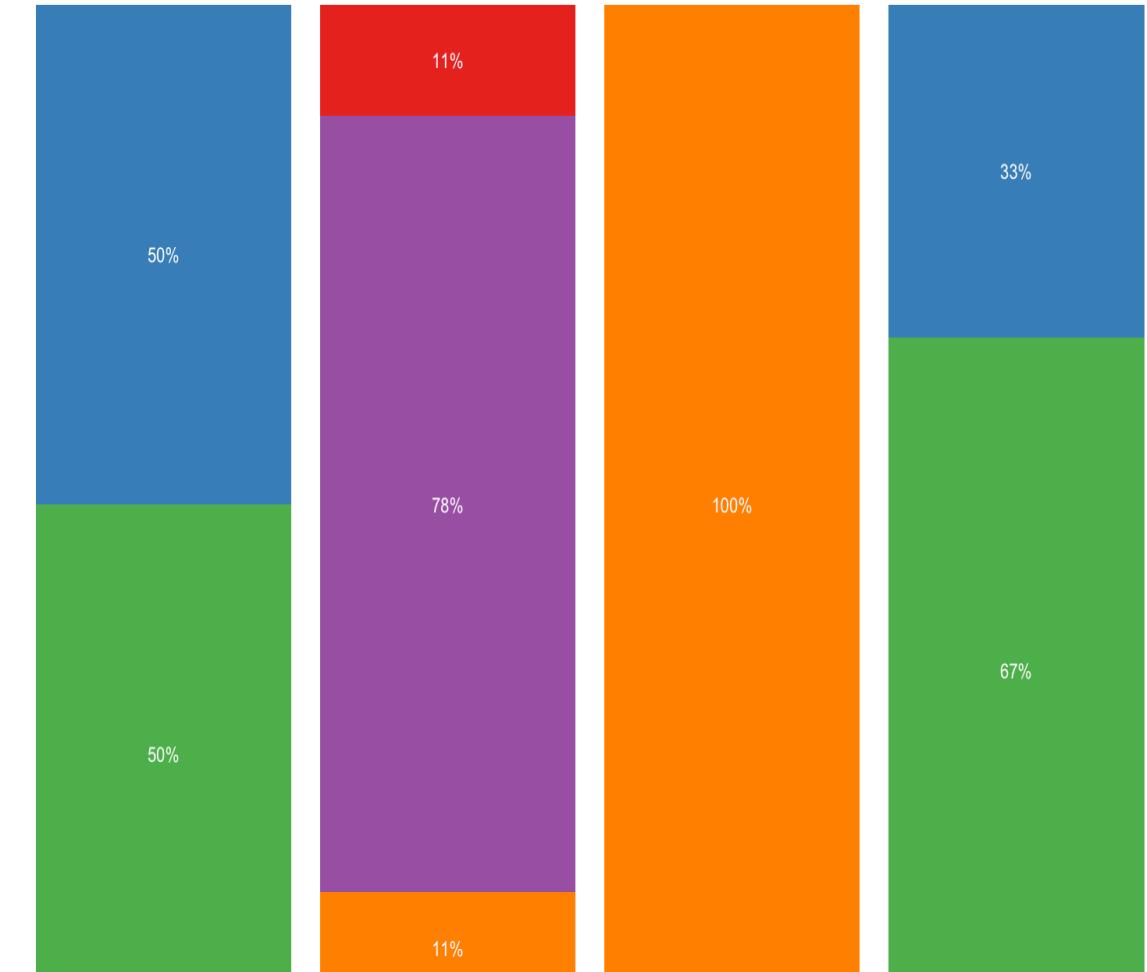
K type Distribution per region



Most prevalent ST : K and O type Distribution

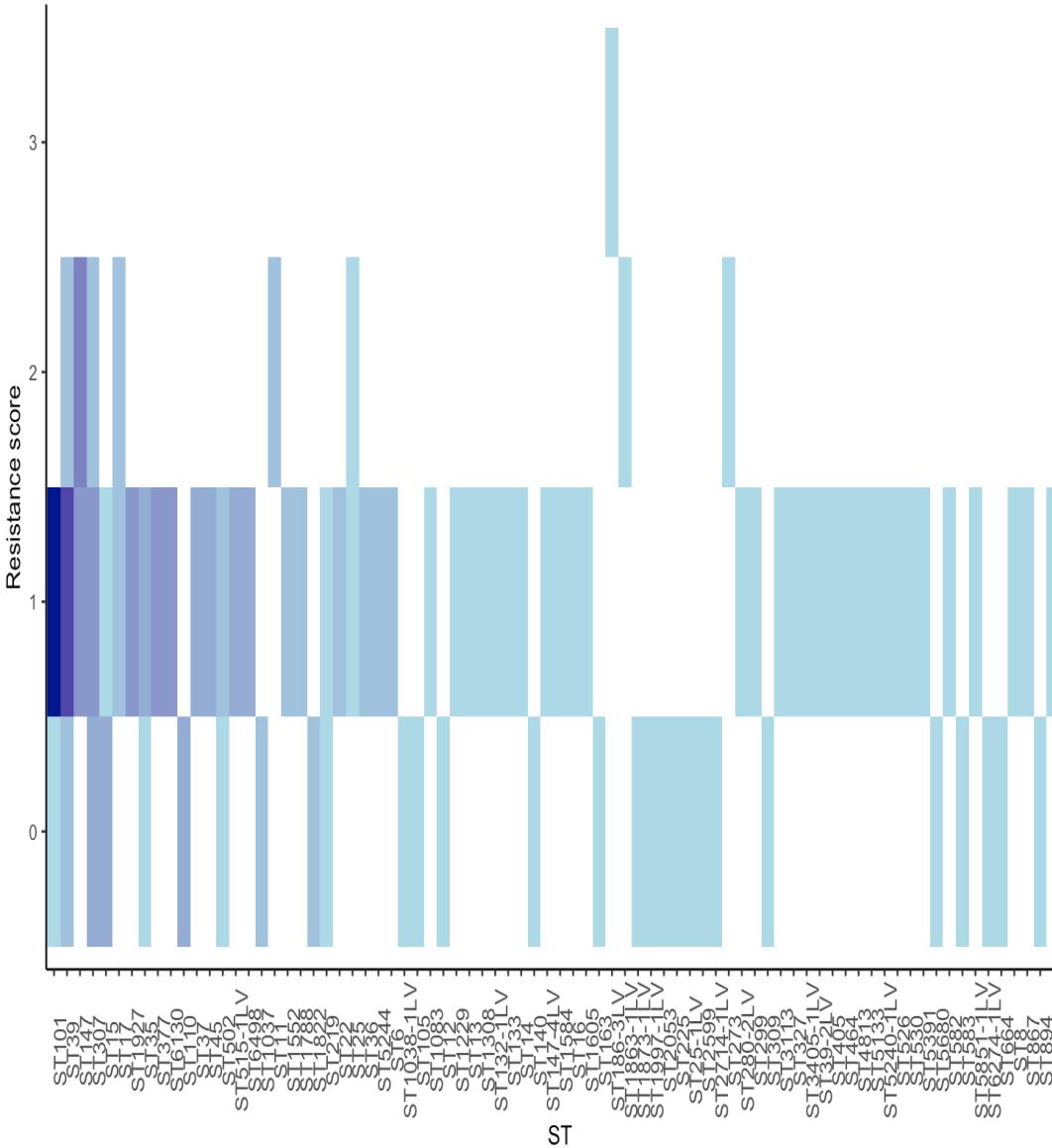


a) Distribution of K type within most prevalent ST

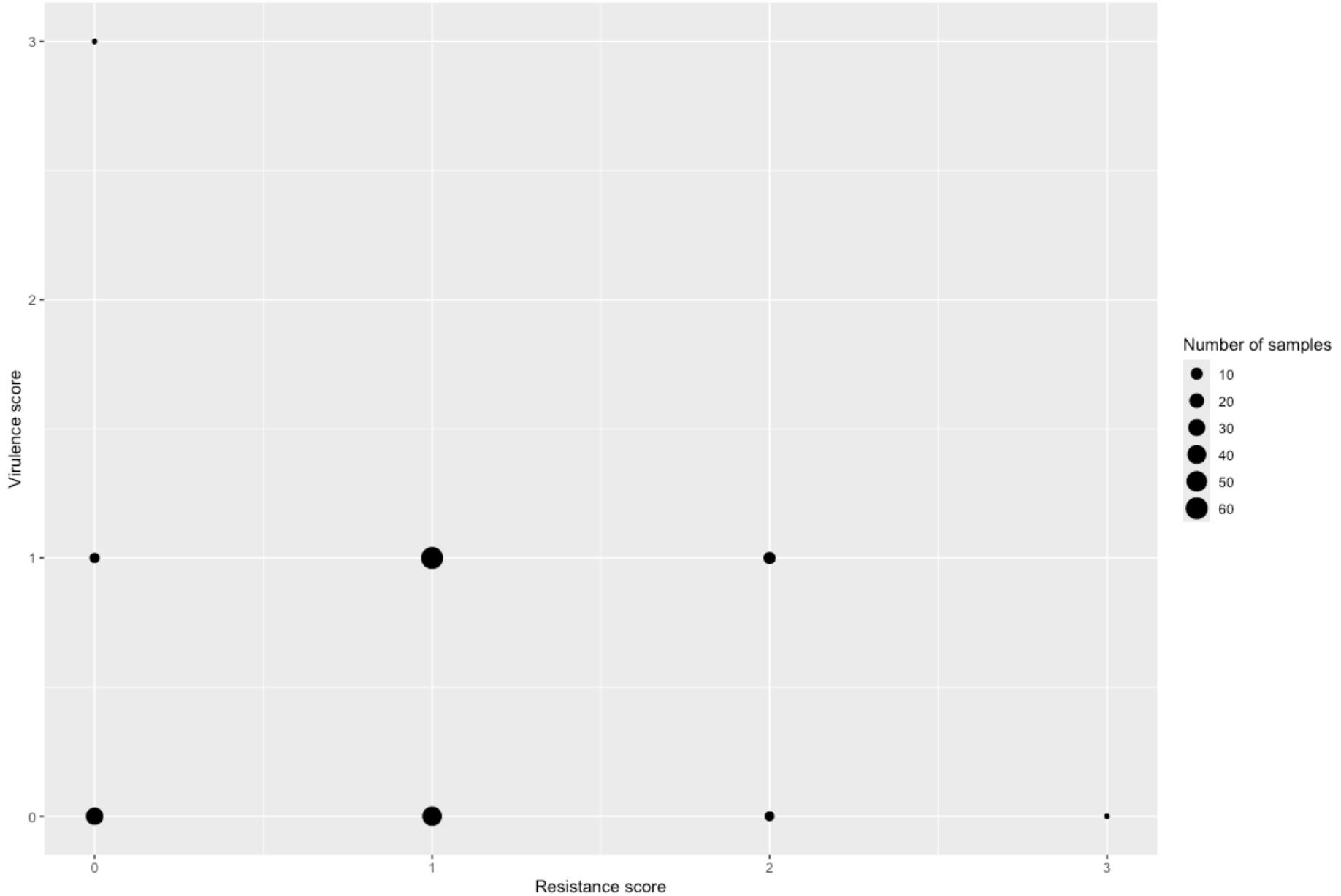


b) Distribution of O type within most prevalent ST

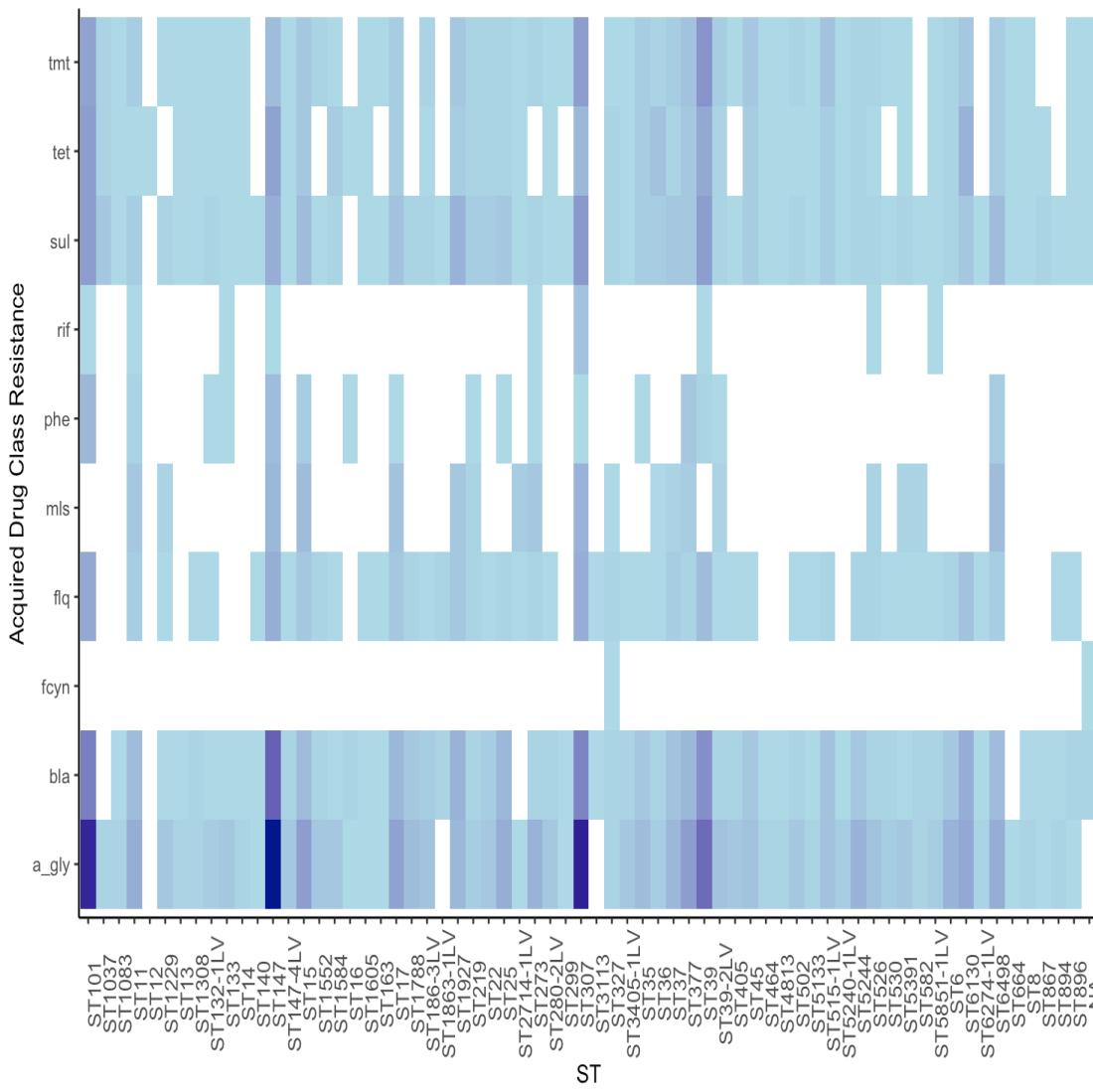
Per ST resistance and virulence score



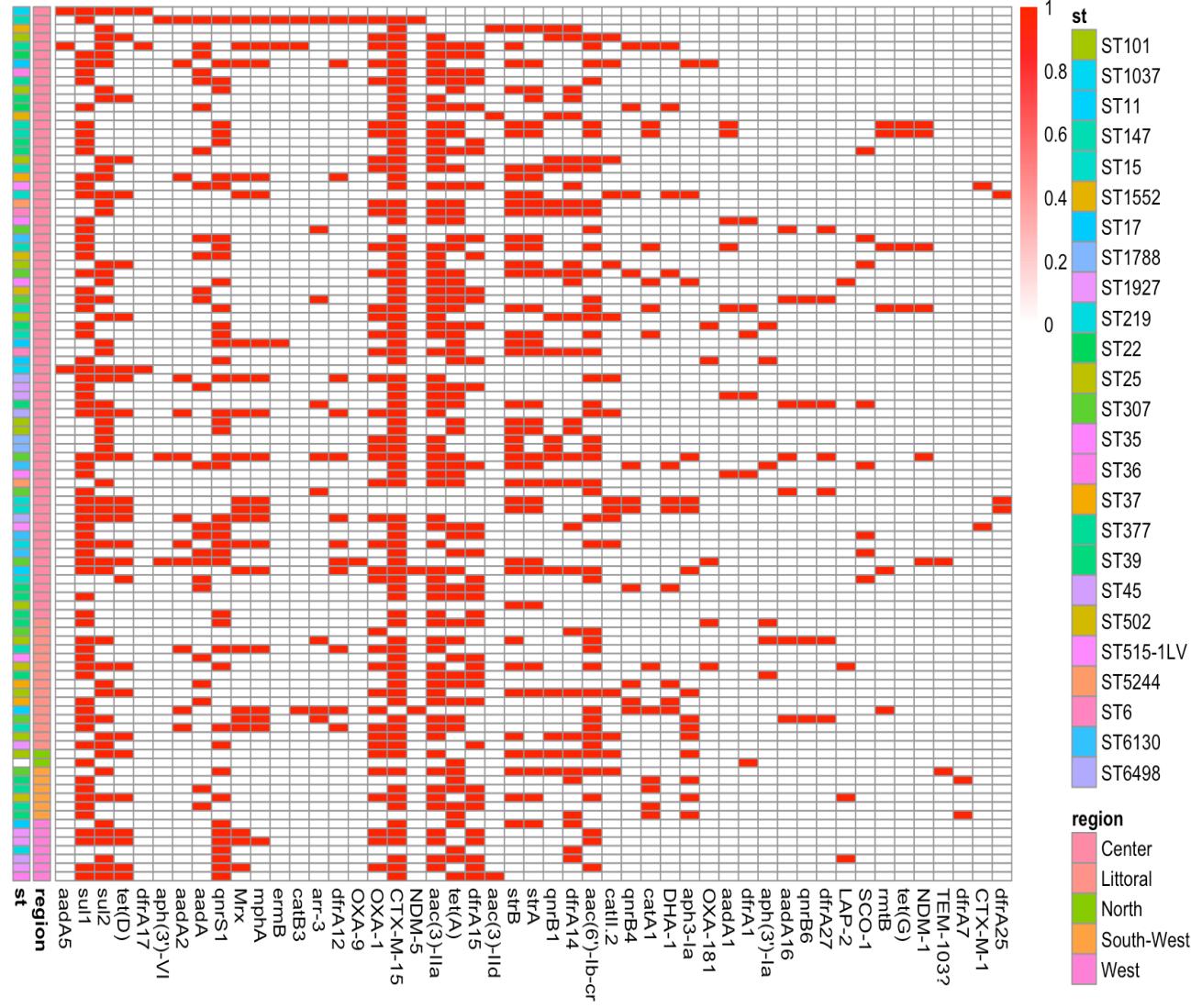
Convergence of resistance and virulence



Most prevalent ST : Resistance Classes and Associated Genes



a) Distribution of acquired drug resistance class per ST

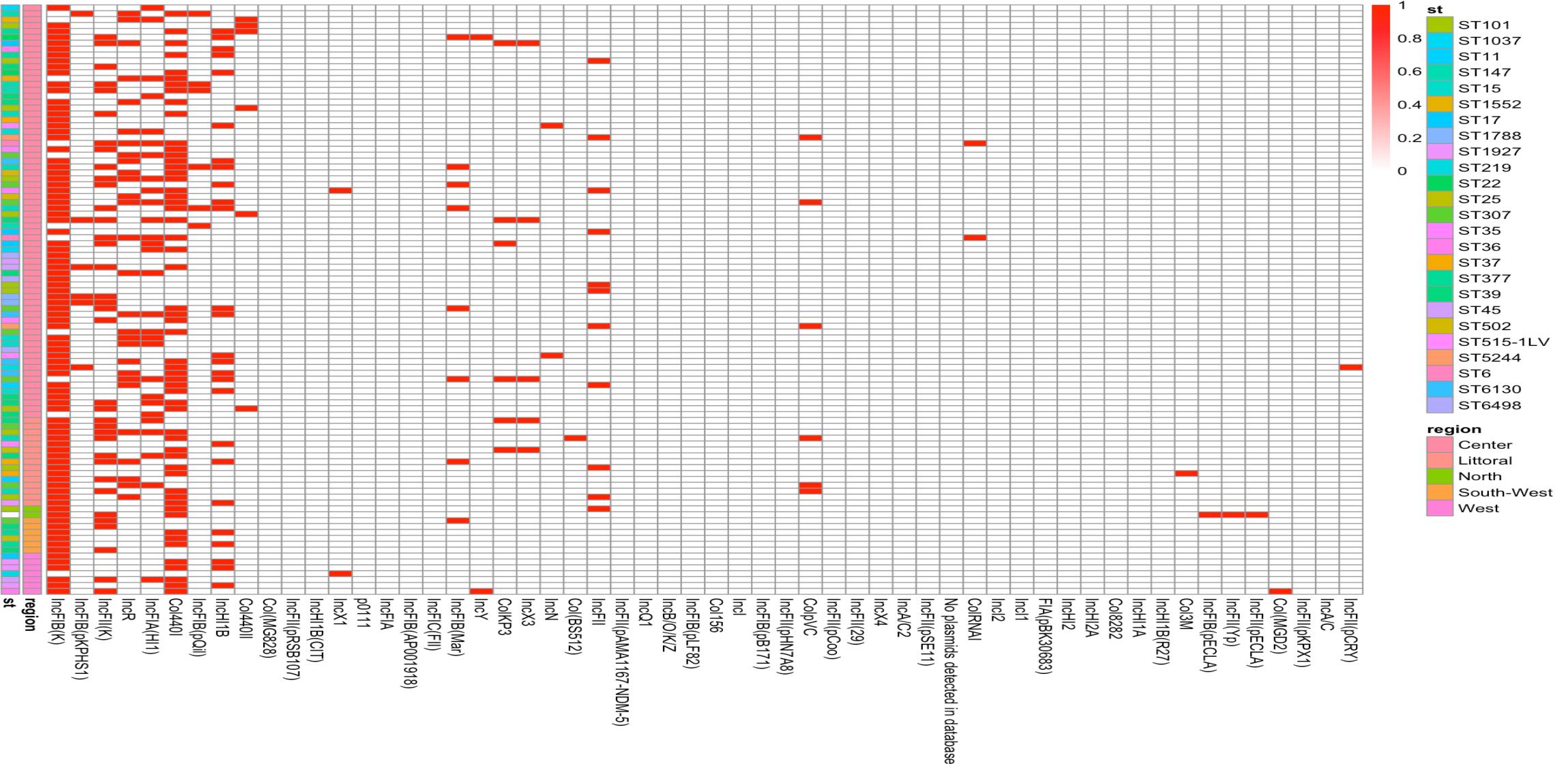


b) Distribution of resistance genes per region and 30 most prevalent ST

These two plots show MDR in our isolates, with **CTX-M-15**, **aac(3)-Ila**, **sul1** and **tet(A)** being the most prevalent resistance genes across the 30 most common STs

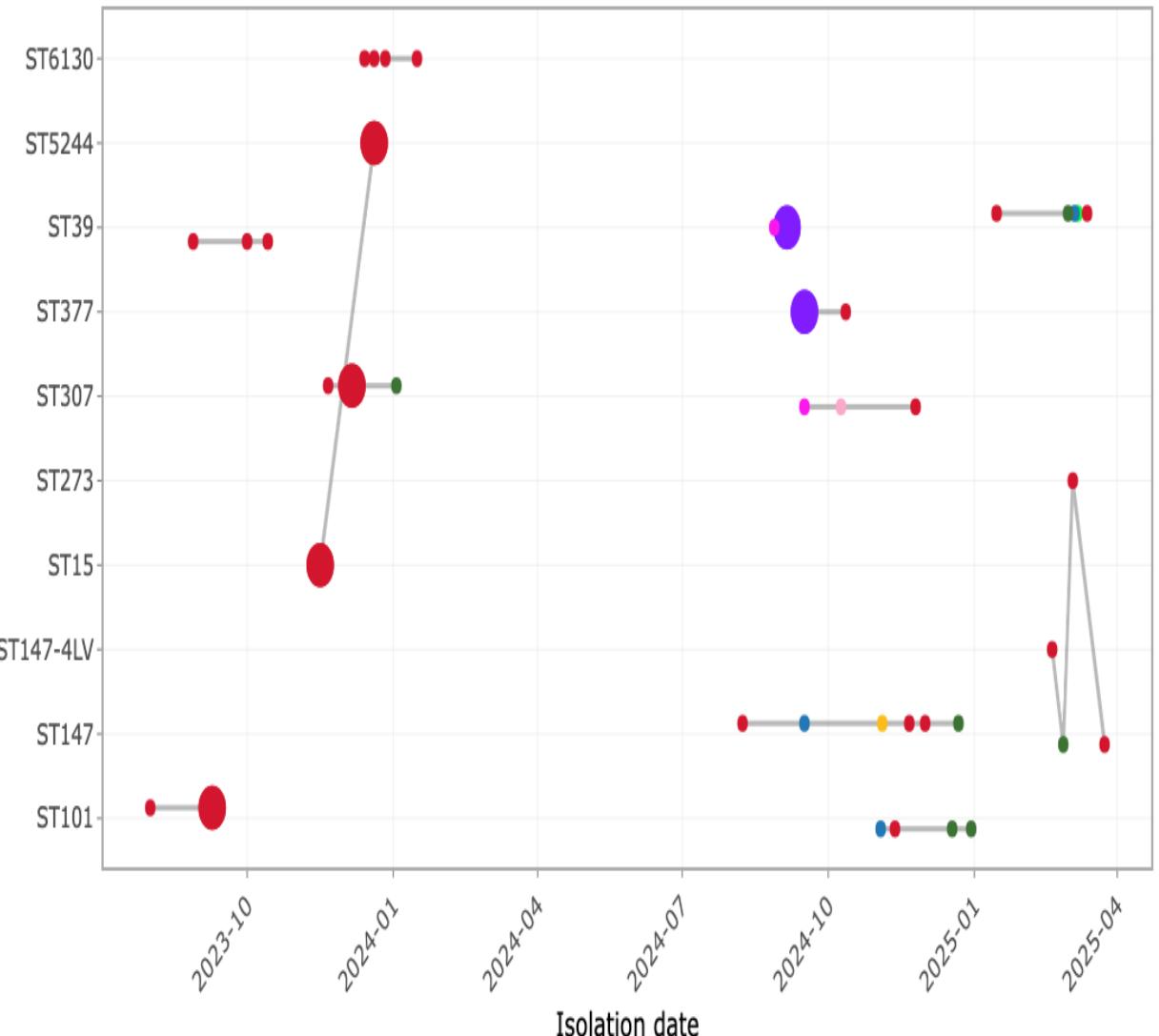


Most prevalent ST : Plasmids distribution



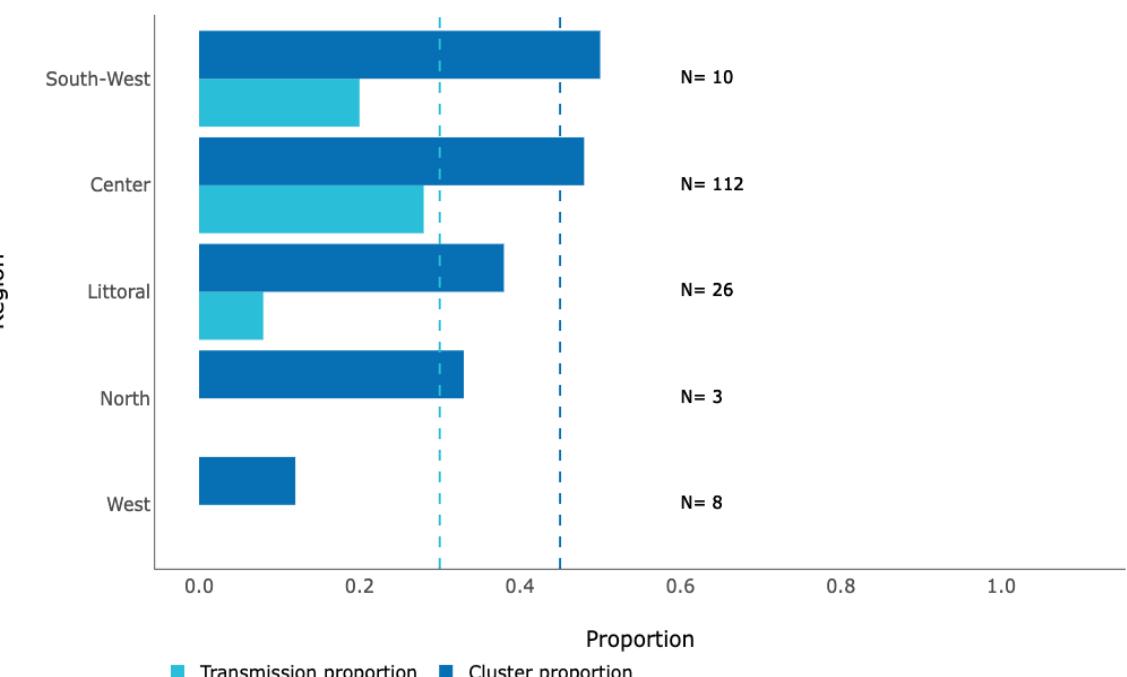
Of the 30 most prevalent STs, only 25 carry plasmids. Among these strains, the majority harbor the *IncFIB(K)* plasmid.

Results of transmission analysis (1/2)



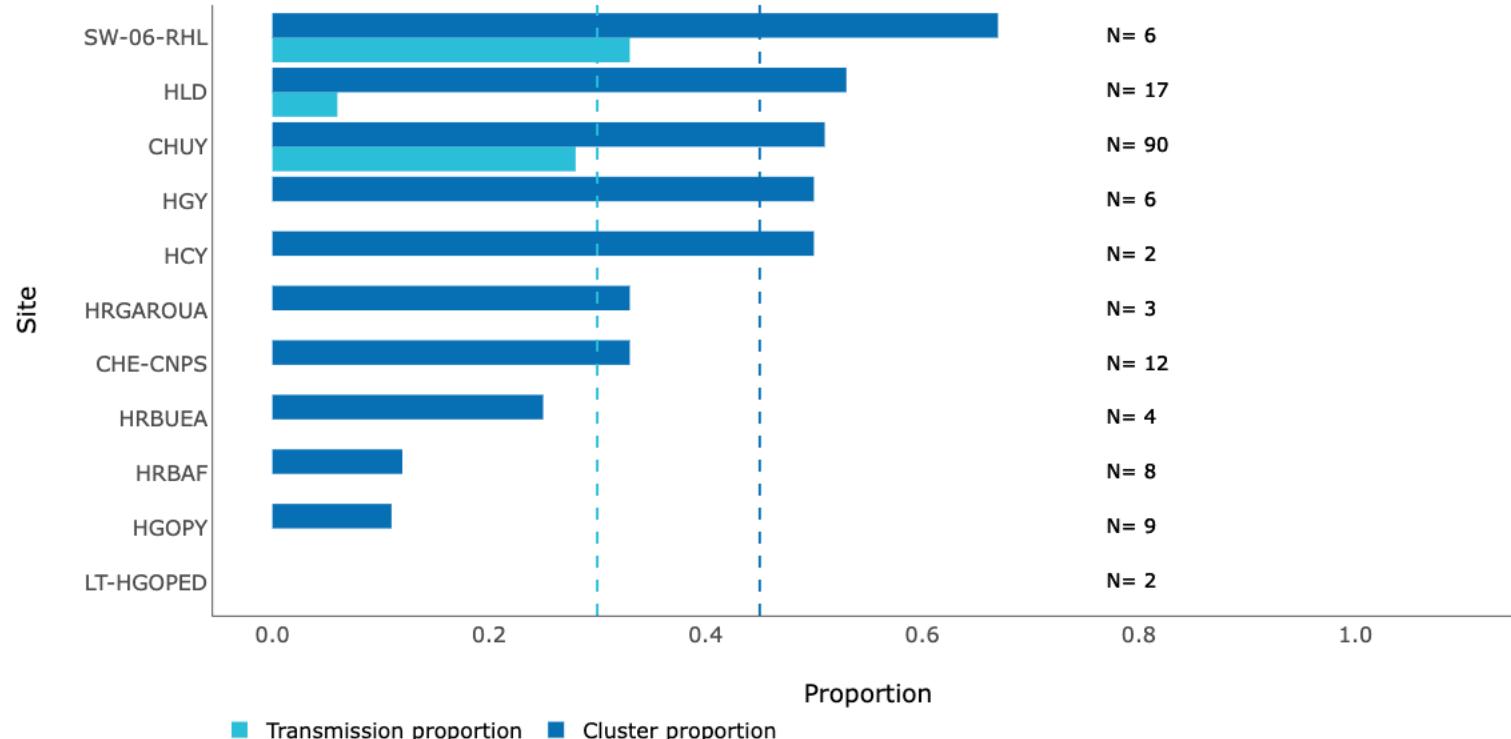
a) Clusters plot per site

- Using hyperparameters :
 - Distance threshold = 30
 - Temporal distance threshold (weeks) = 7
 - Spatial Cluster = Country
- There are 12 clusters in our data with a country clustering
- Observed transmissions:
 - Between laboratories in the Centre and Littoral regions
 - Between laboratories in the Centre and South-West regions
 - Between laboratories within the Centre region
- The Center region is the one with the most transmission



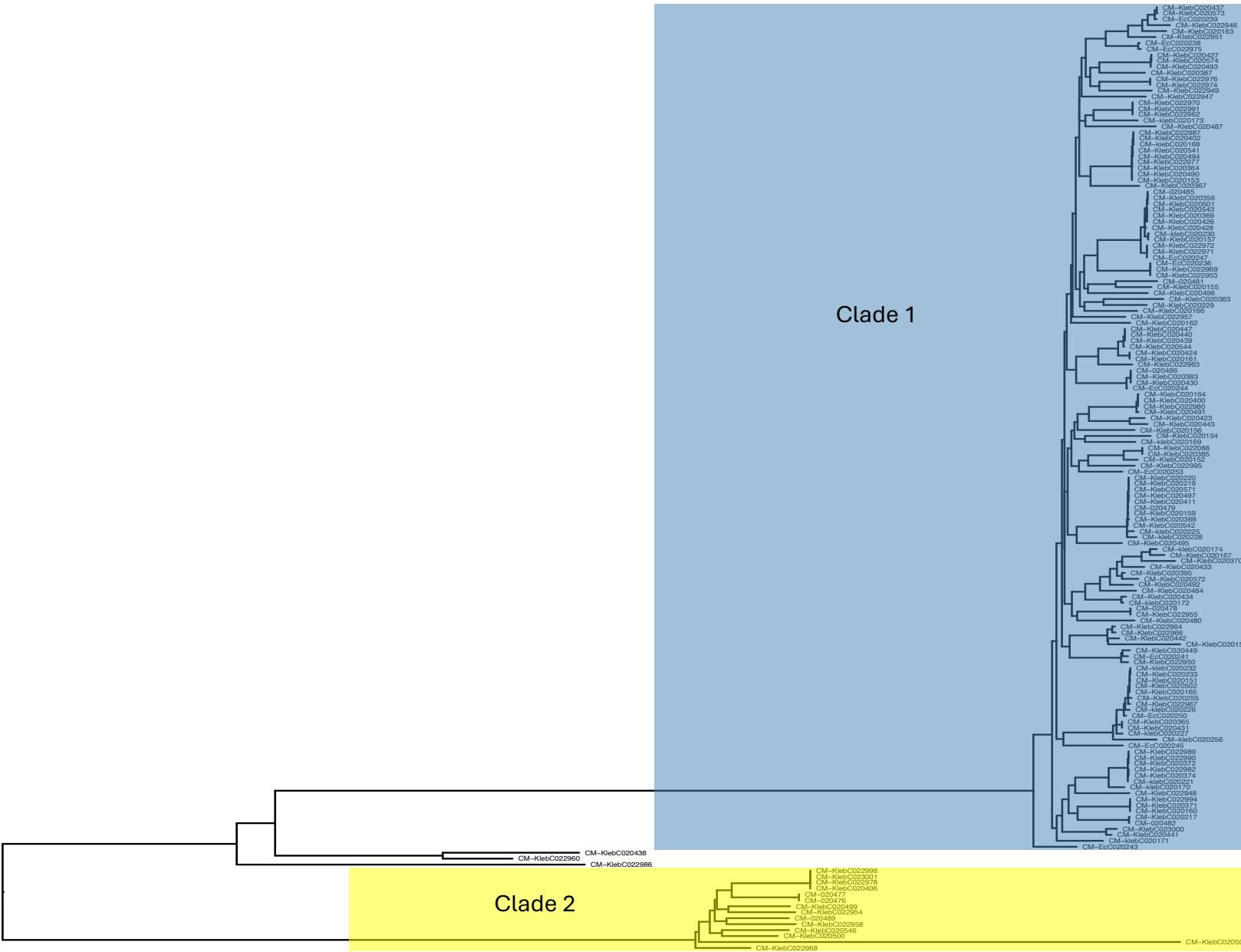
b) Stratified cluster statistics

Results of transmission analysis (2/2)

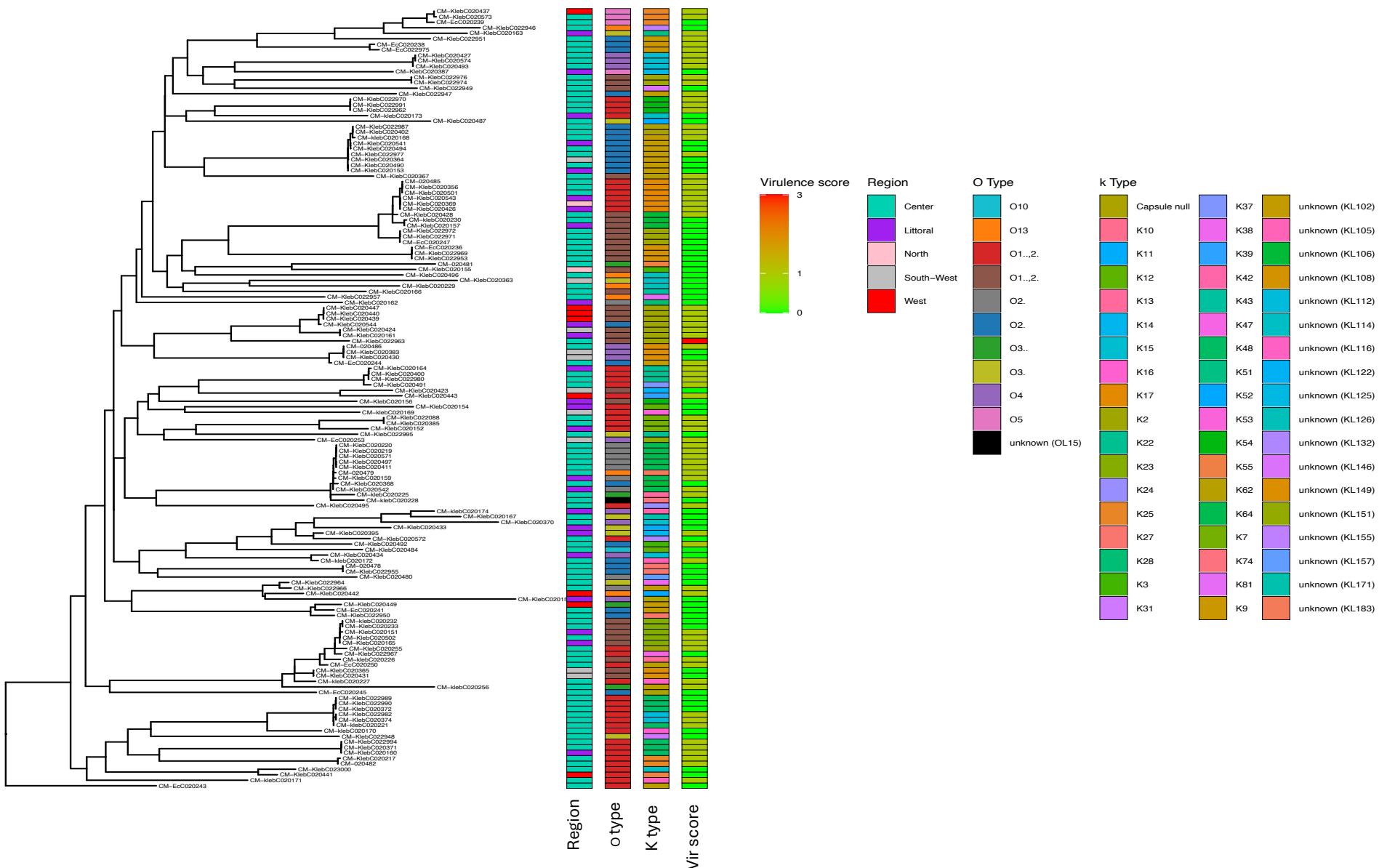


- The SW-06-RHL is the laboratory with the highest number of intrahospital transmissions, followed by CHUY and then HLD.
- No intrahospital transmissions were observed in the other laboratories.

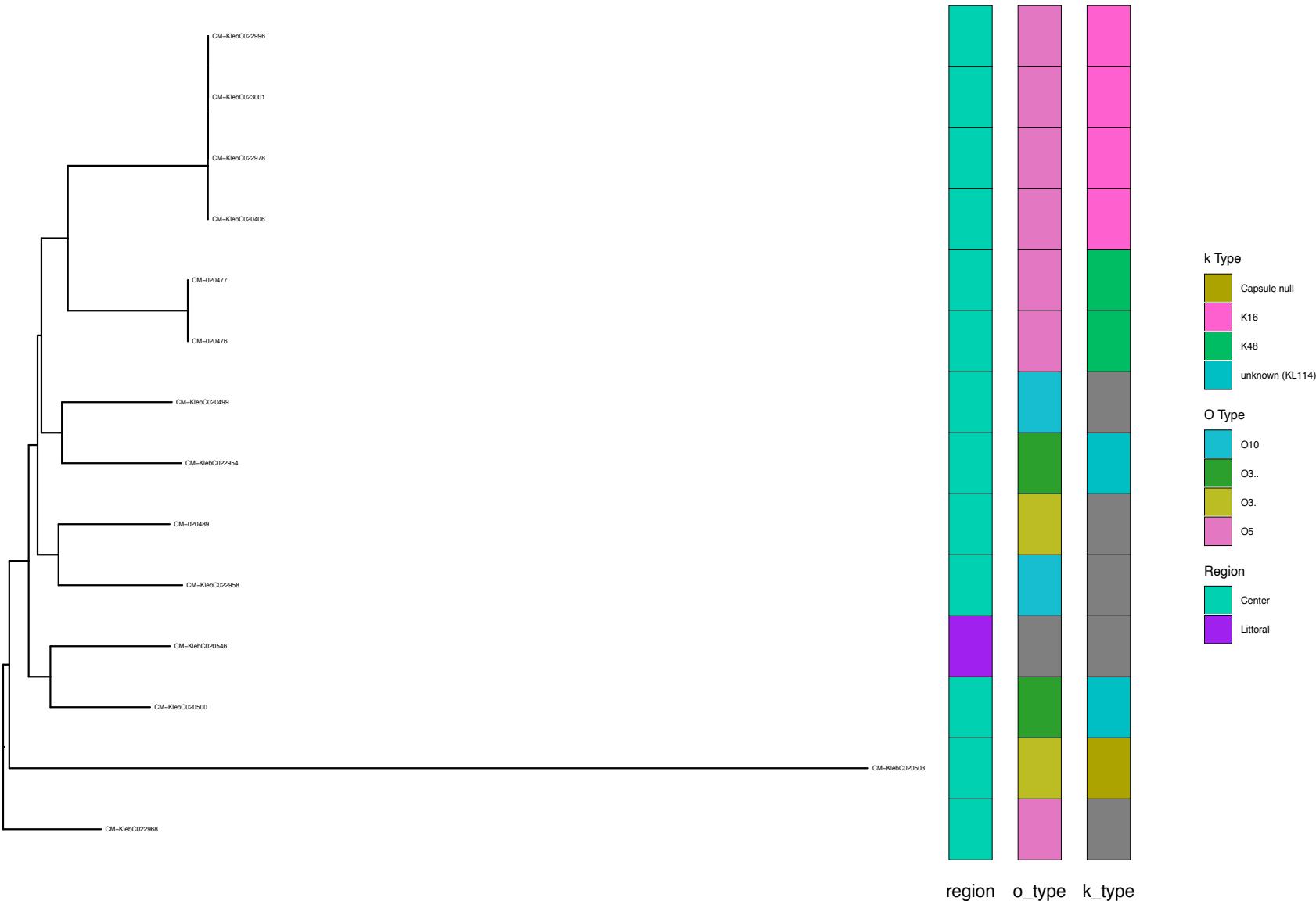
Results of phylogenetic analysis (1/3)



Results of phylogenetic analysis : Clade 1 (2/3)



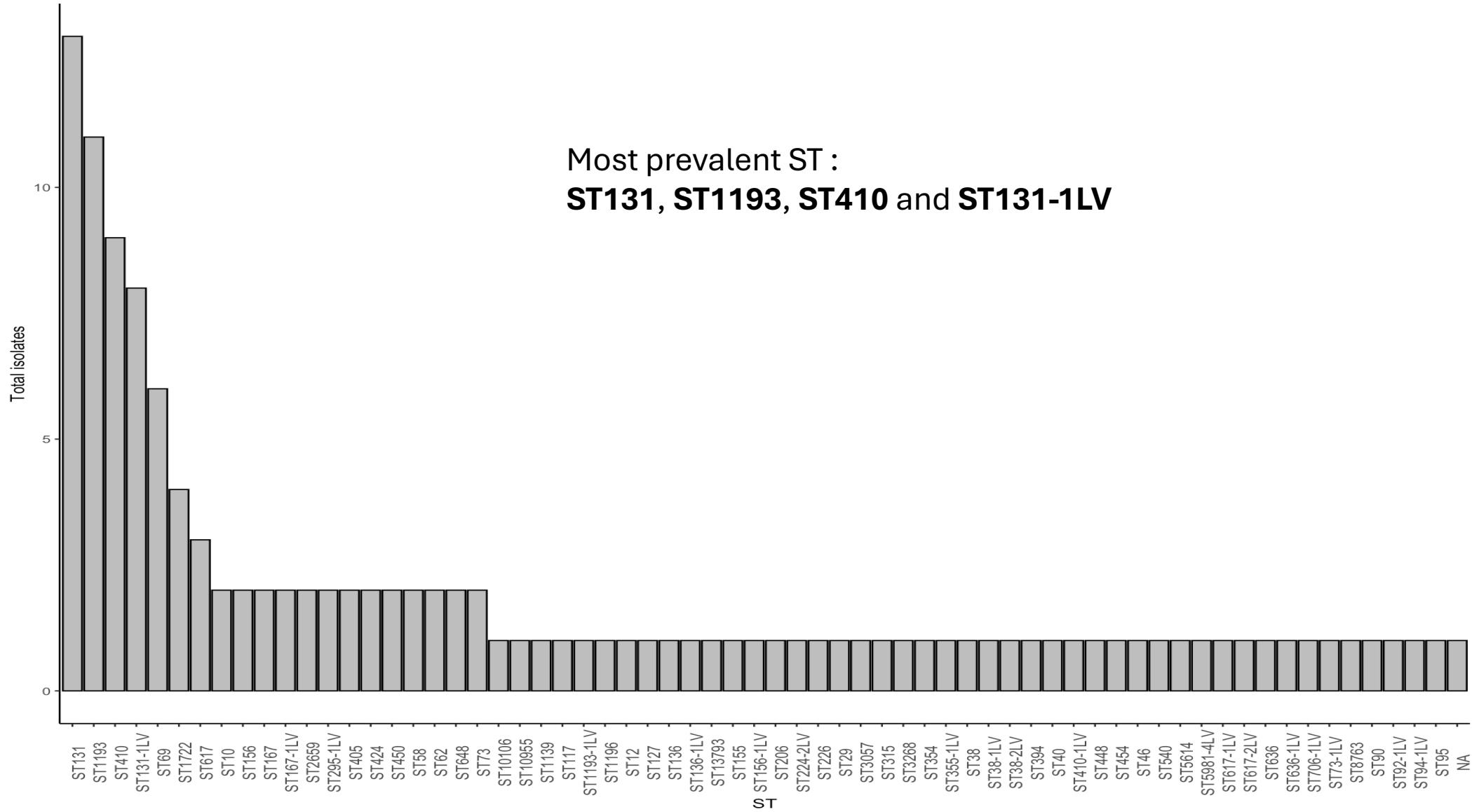
Results of phylogenetic analysis : Clade 2 (3/3)



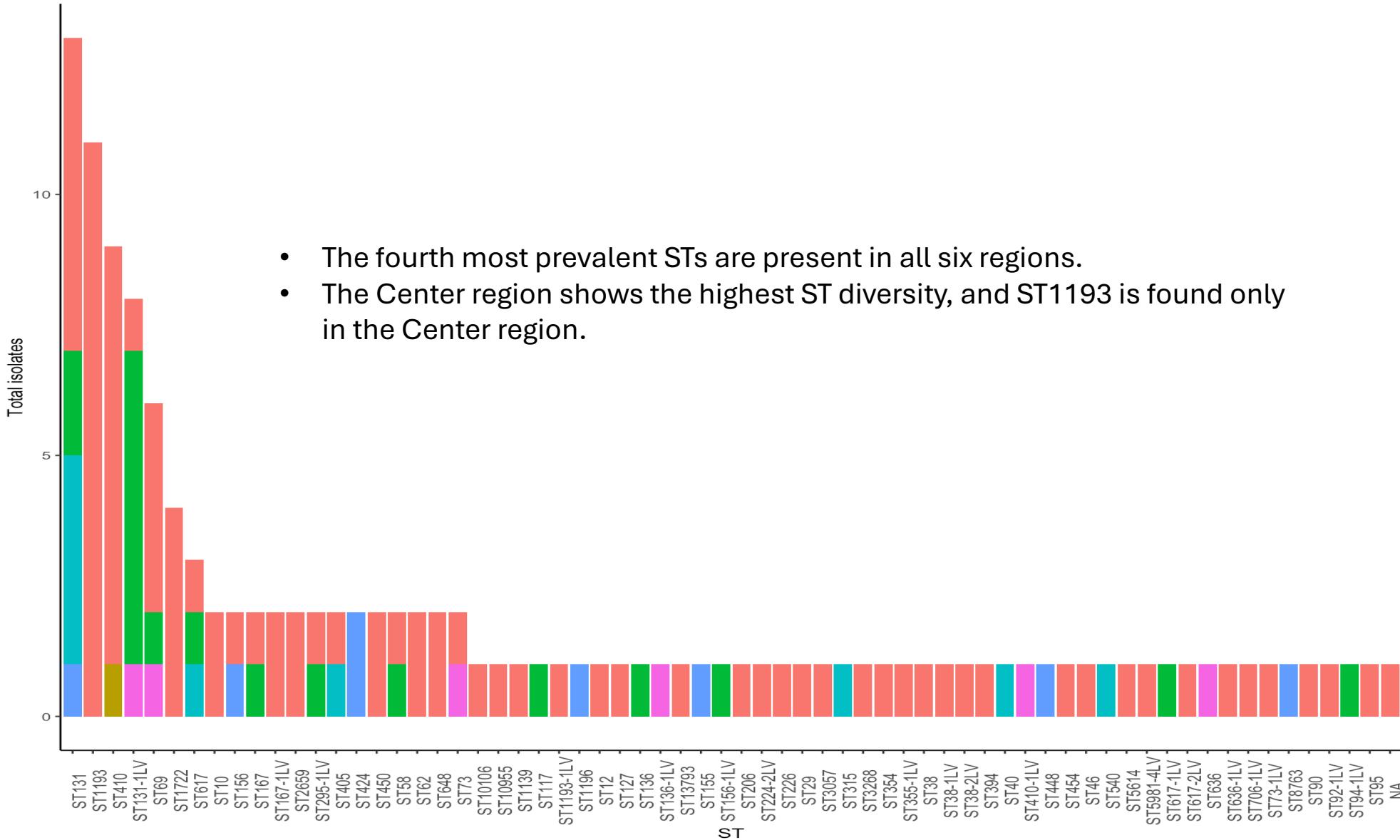


E. COLI/SHIGELLA ANALYSIS RESULTS

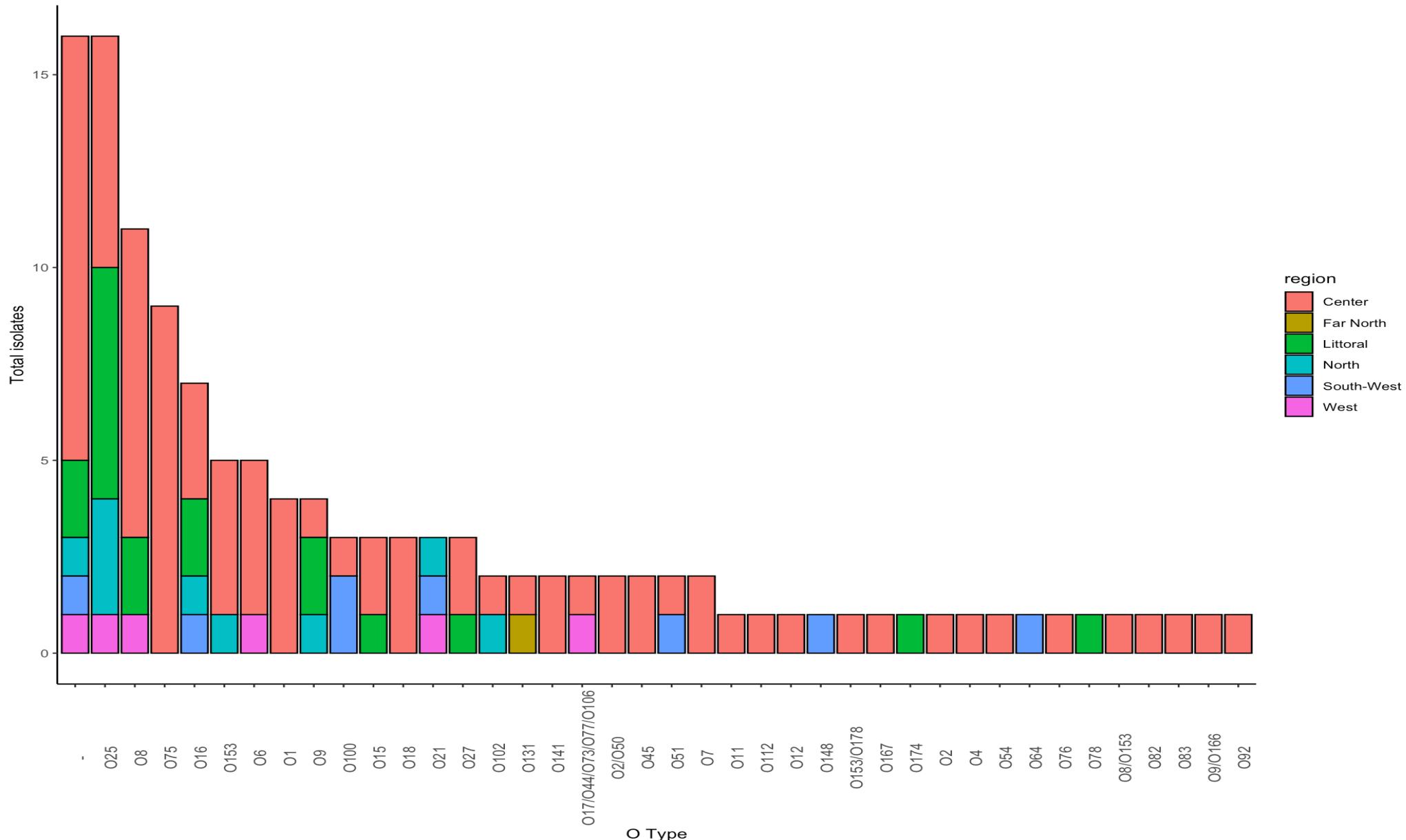
ST prevalence distribution



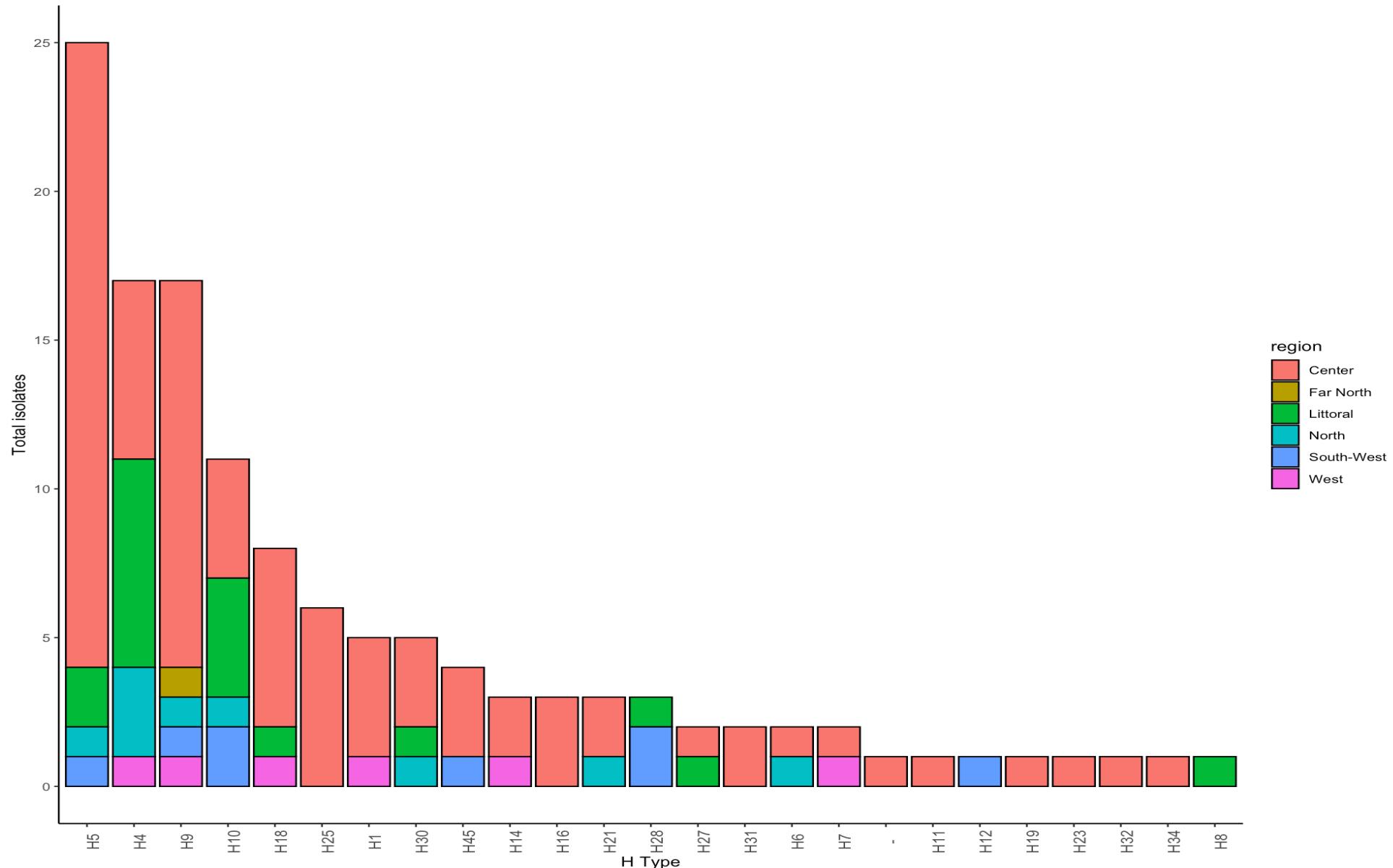
ST prevalence distribution per region



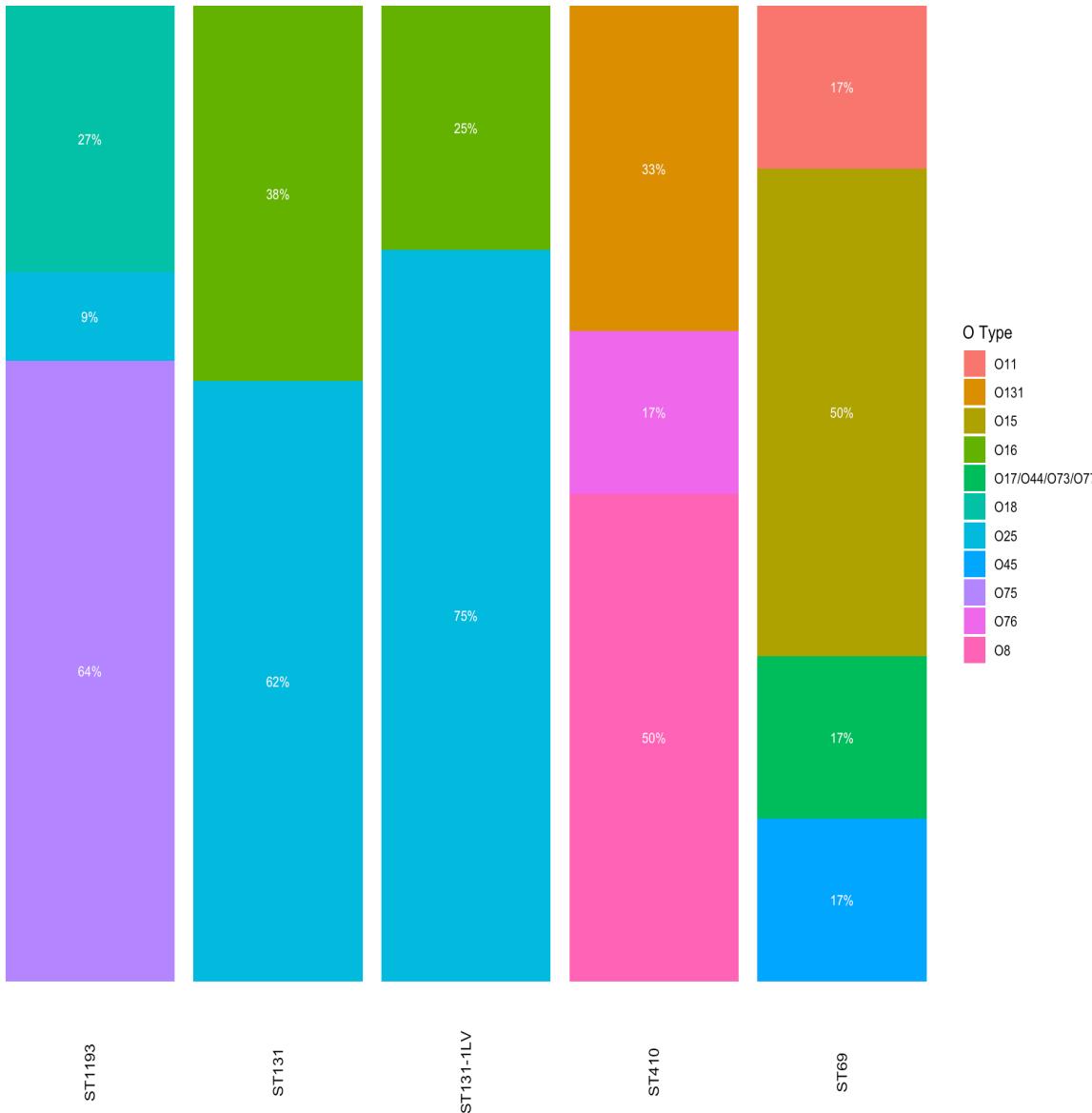
O type Distribution per region



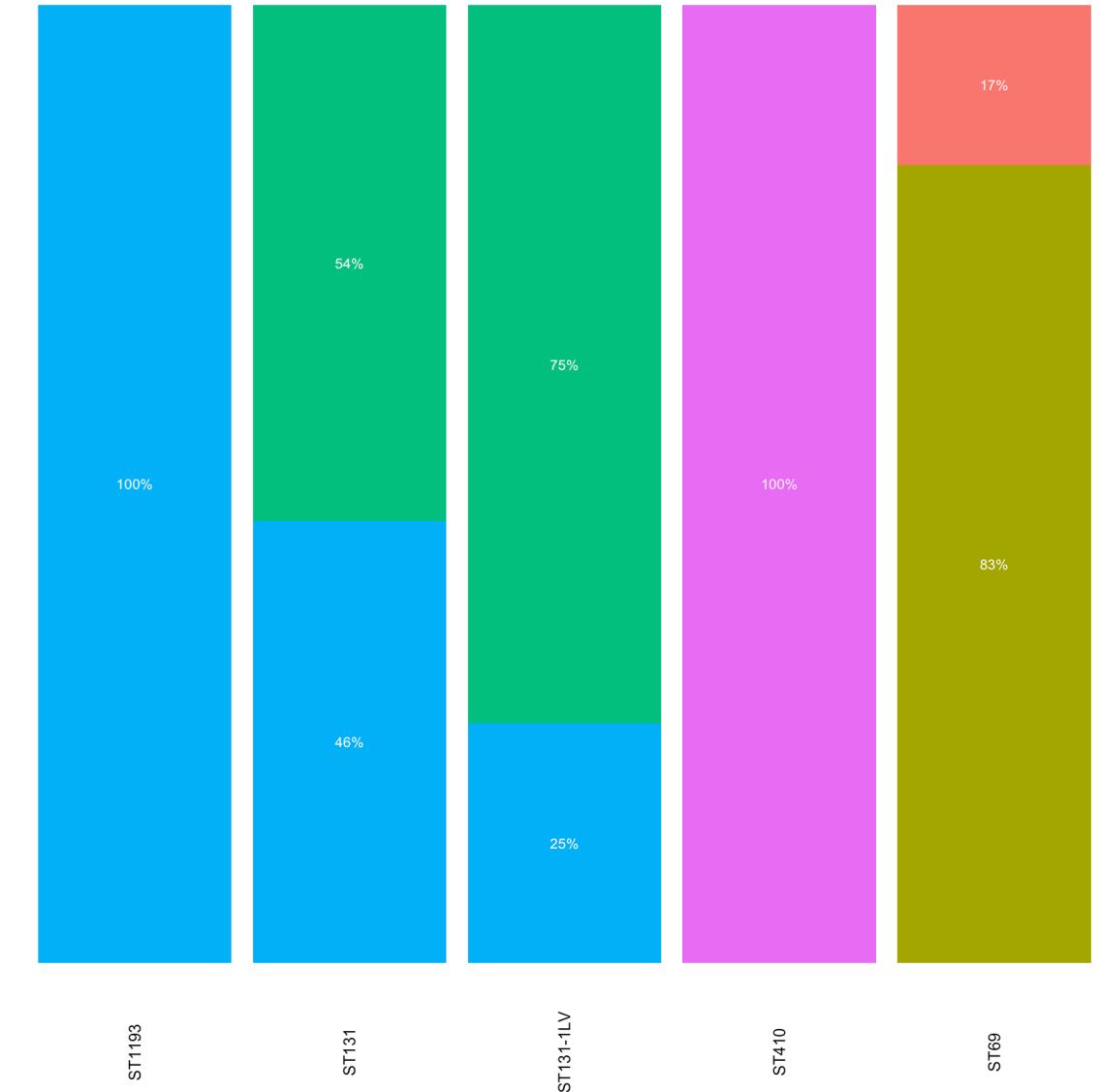
H type Distribution per region



5 most prevalent ST : O type and H type distribution

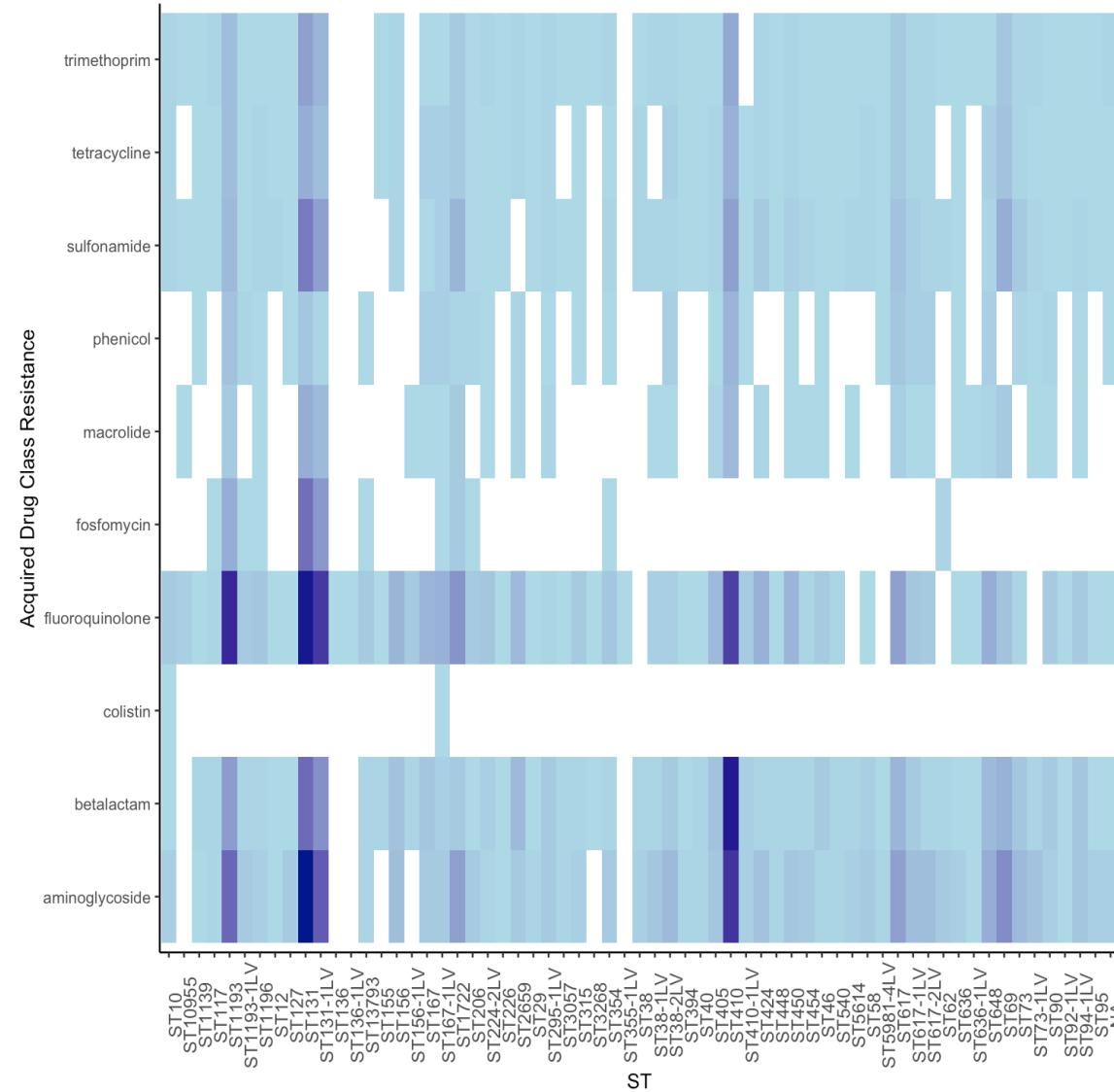


a) Distribution of O type within most prevalent ST

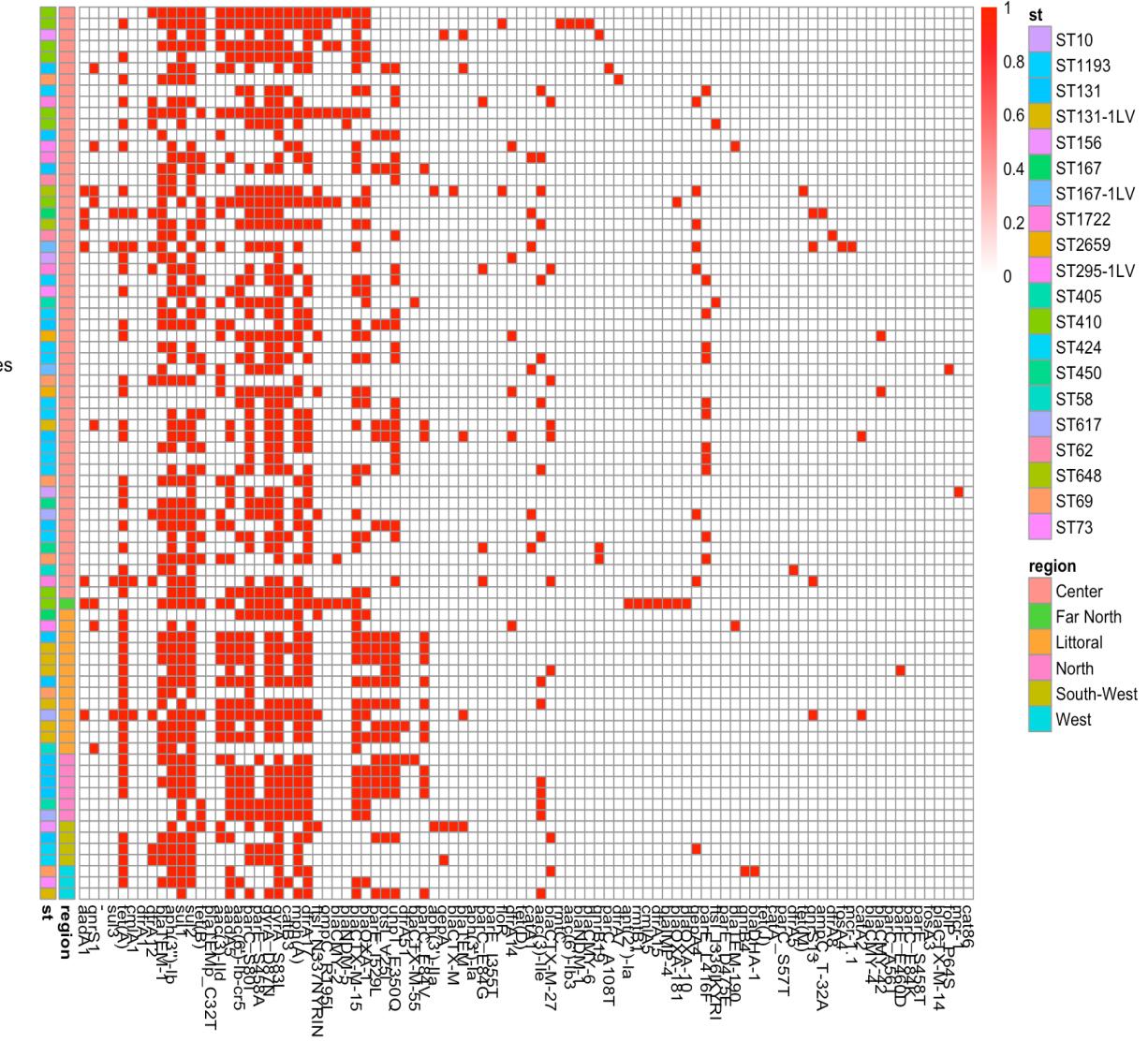


b) Distribution of H type within most prevalent ST

Most prevalent ST: Resistance Classes and Associated Genes



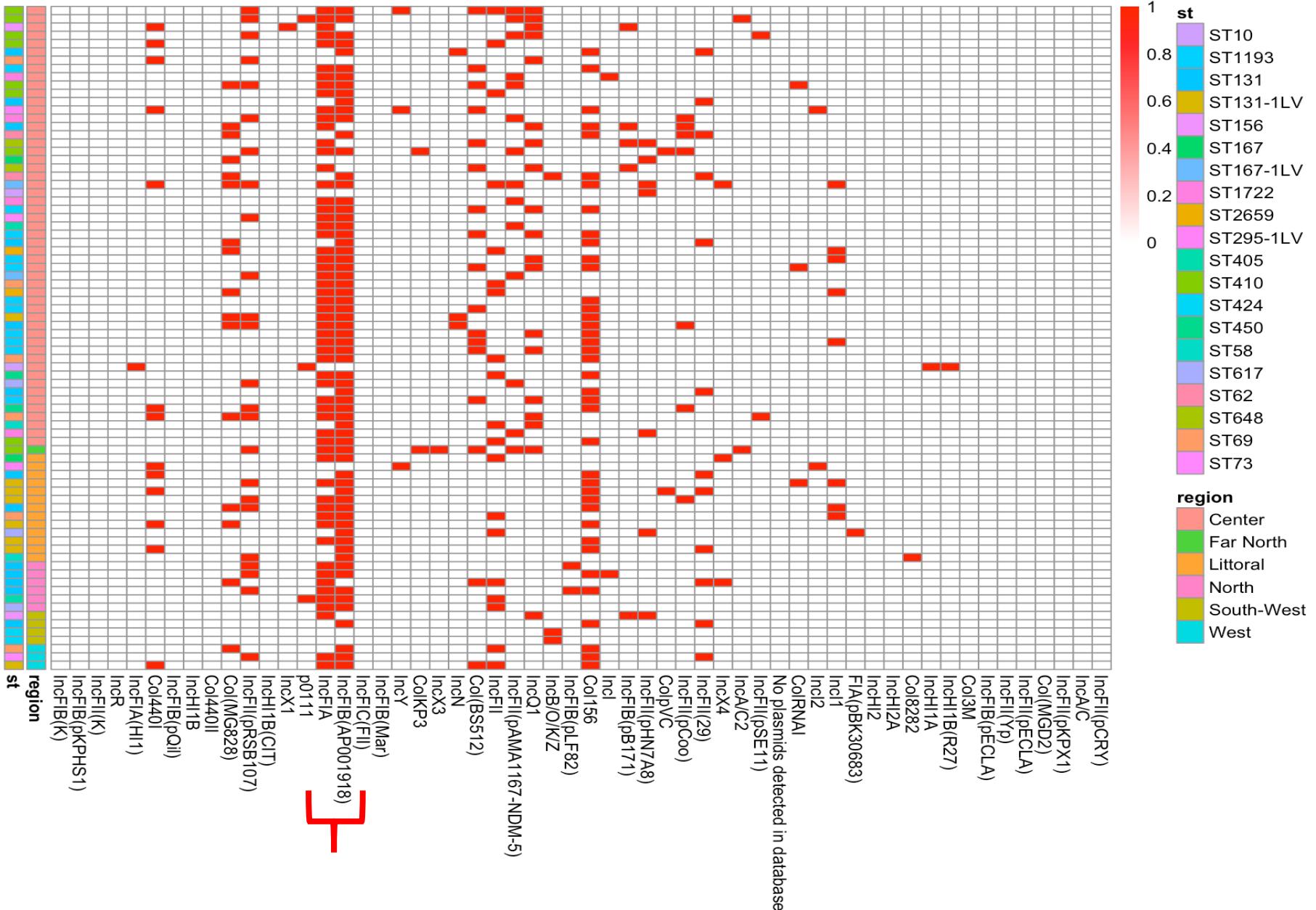
a) Distribution of acquired drug resistance class per ST



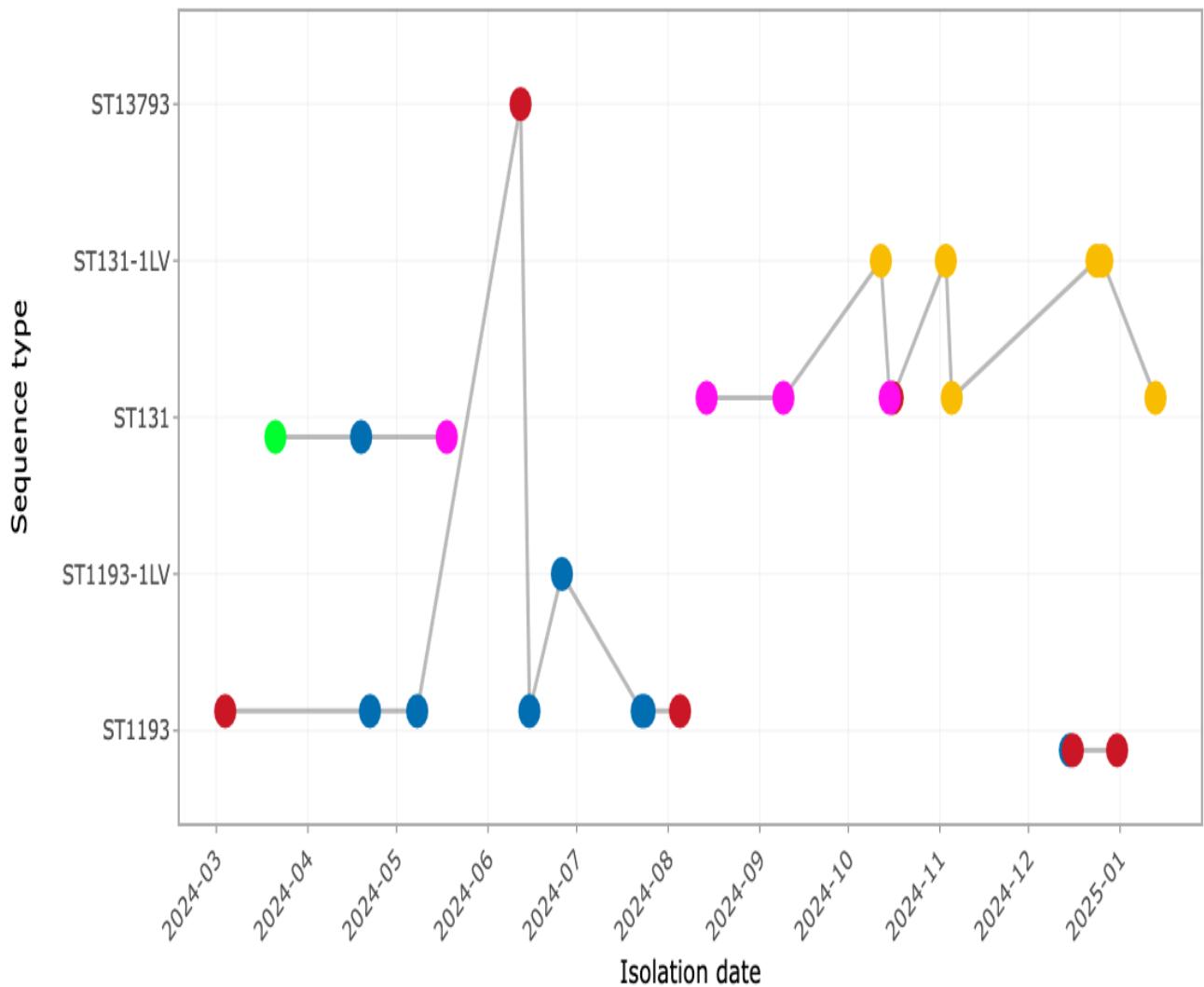
b) Distribution of resistance genes per region and 20 most prevalent ST



Most prevalent ST : Plasmids distribution

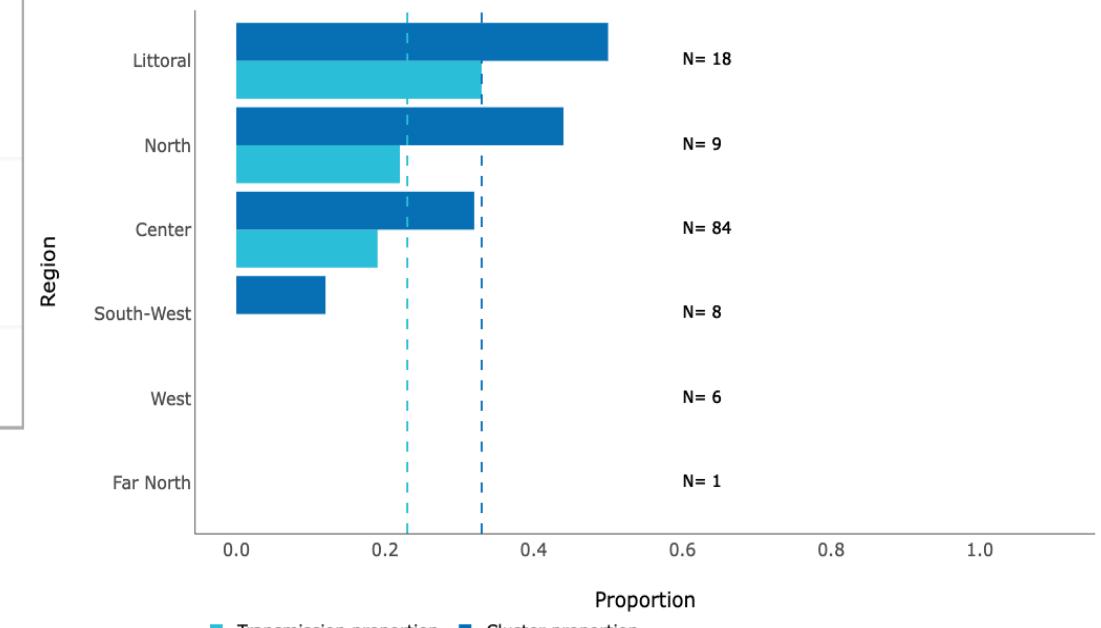


Results of transmission analysis (1/2)



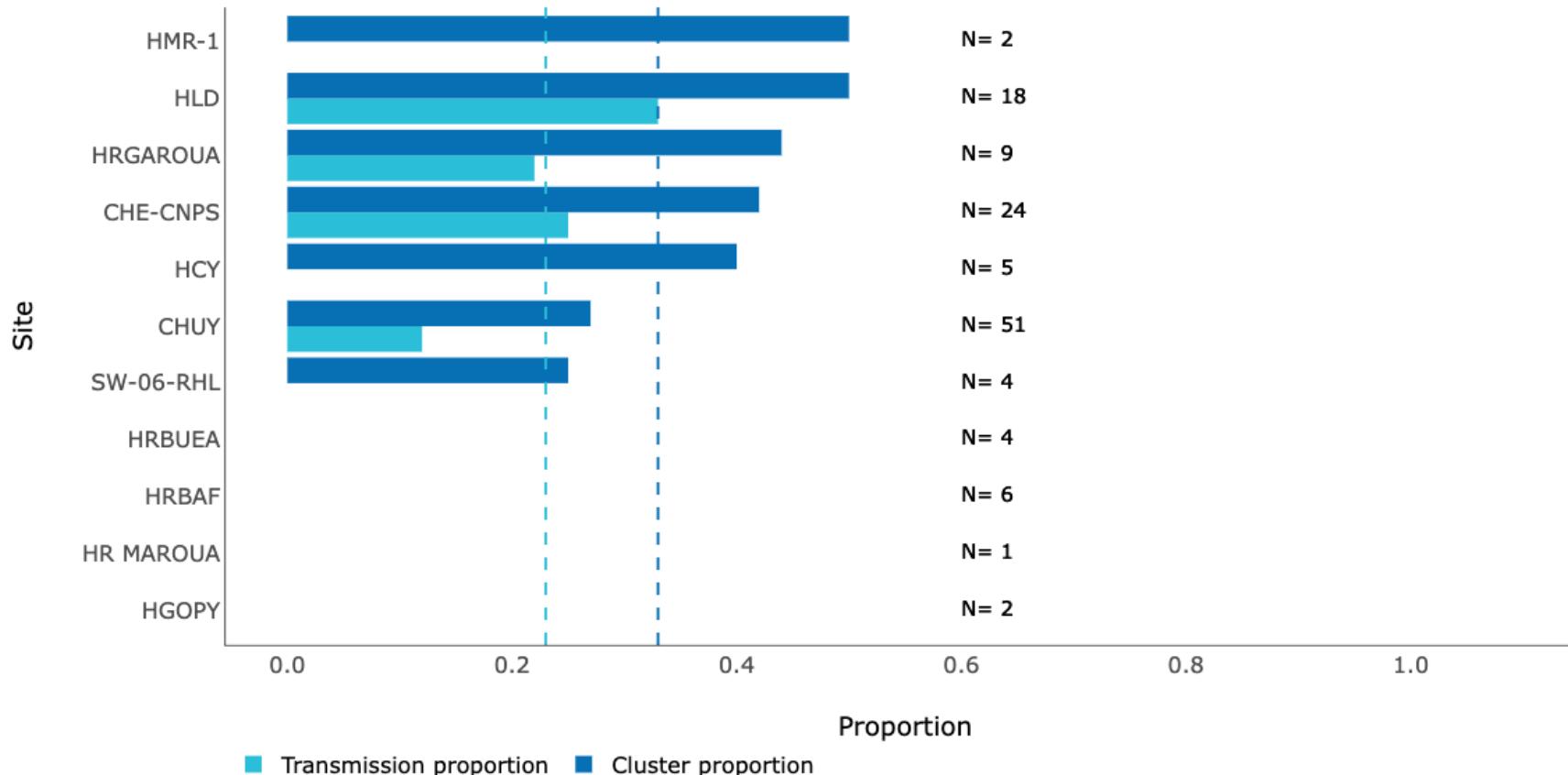
a) Clusters plot per site

- **Using hyperparameters :**
 - Distance threshold = 30
 - Temporal distance threshold (weeks) = 4
 - Spatial Cluster = Country
- There are 4 clusters in our data with a country clustering
- **Observed transmissions:**
 - Between laboratories in the Centre, North and Littoral regions
 - Between laboratories in the Centre and North regions
 - Between laboratories within the Centre region
- **The Littoral region is the one with the most transmission**



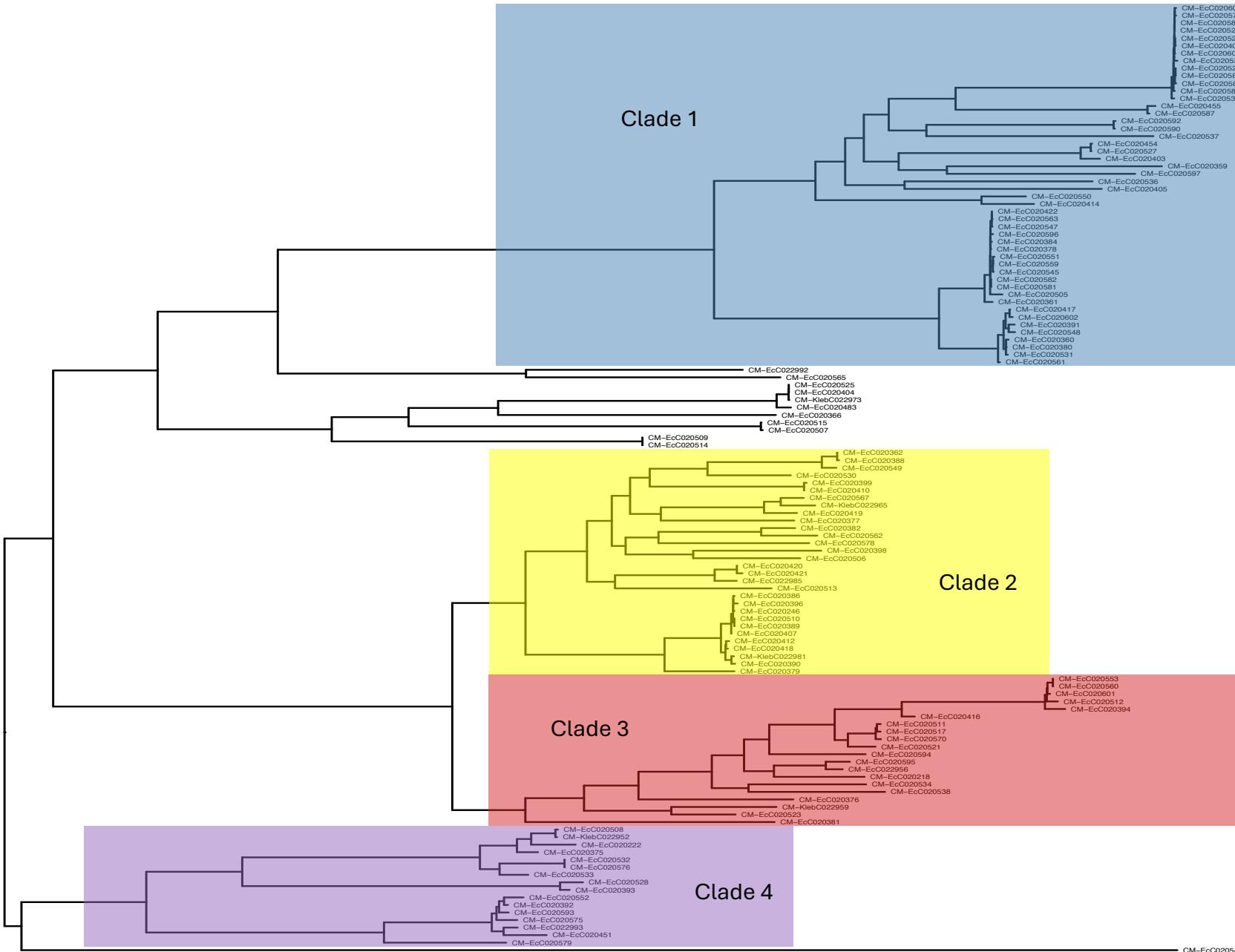
b) Stratified cluster statistics

Results of transmission analysis (2/2)

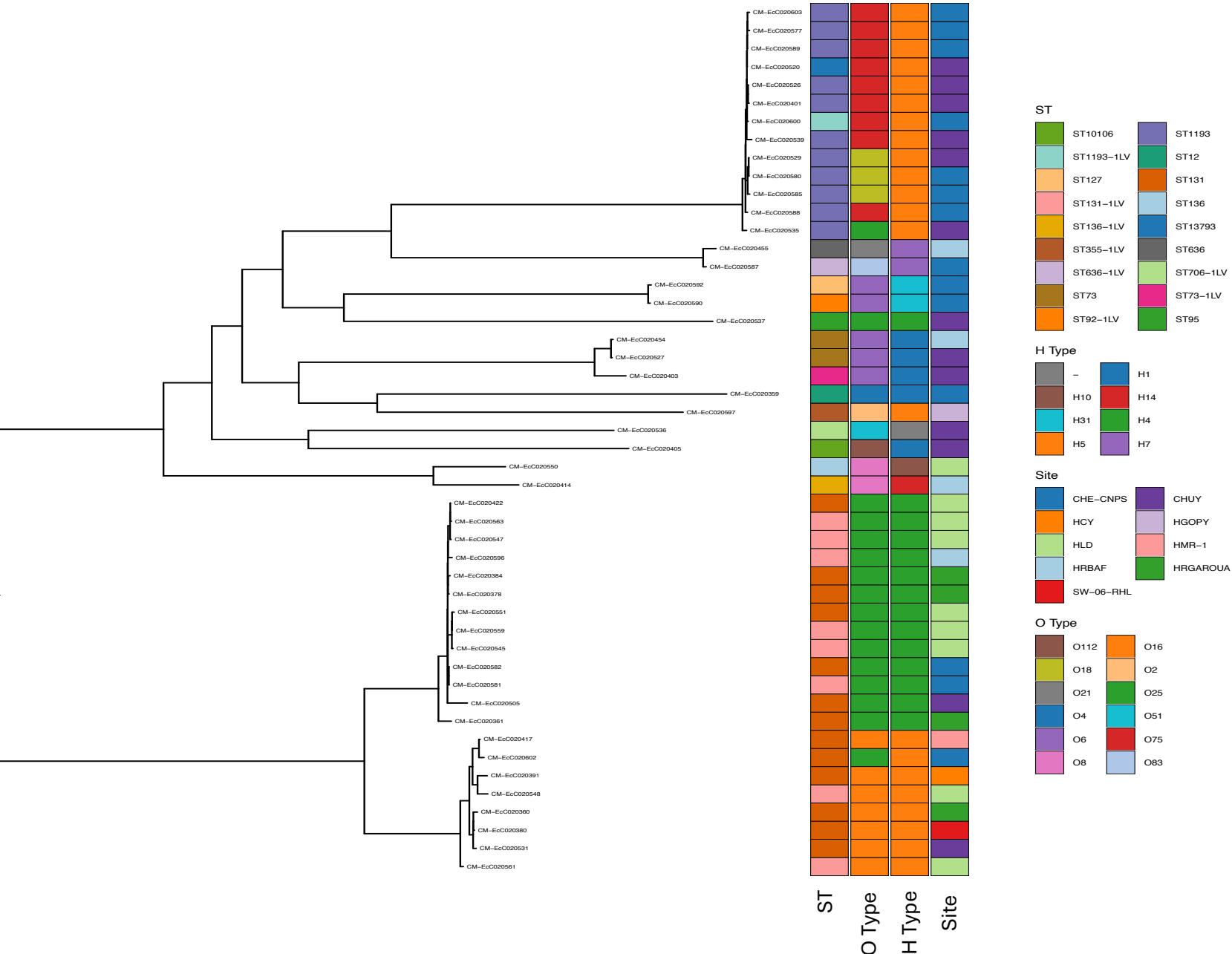


- The HLD is the laboratory with the highest number of intrahospital transmissions, followed by CHE-CNPS, then HRGAROUA and CHUY.
- No intrahospital transmissions were observed in the other laboratories.

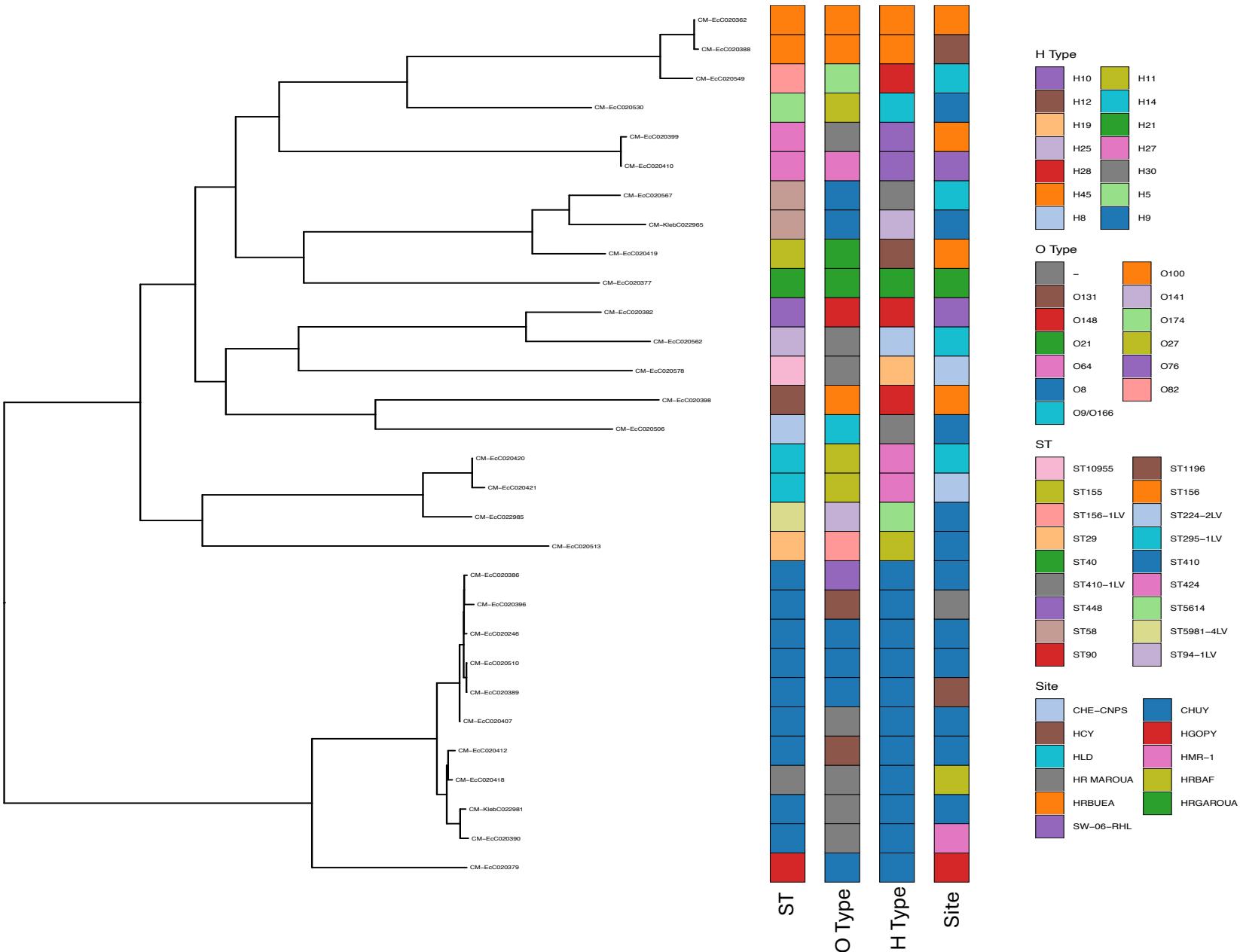
Results of phylogenetic analysis (1/5)



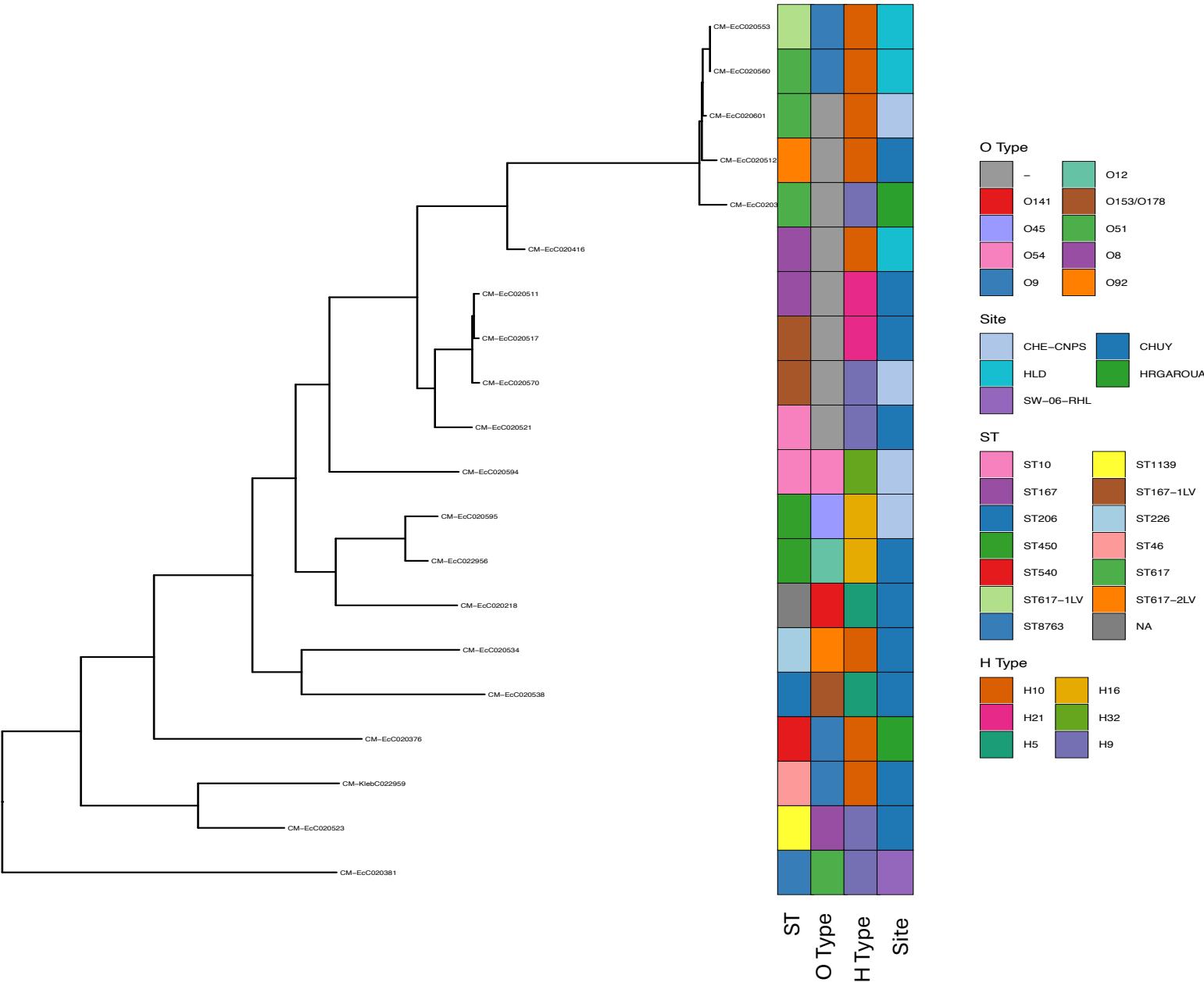
Results of phylogenetic analysis : Clade 1 (2/5)



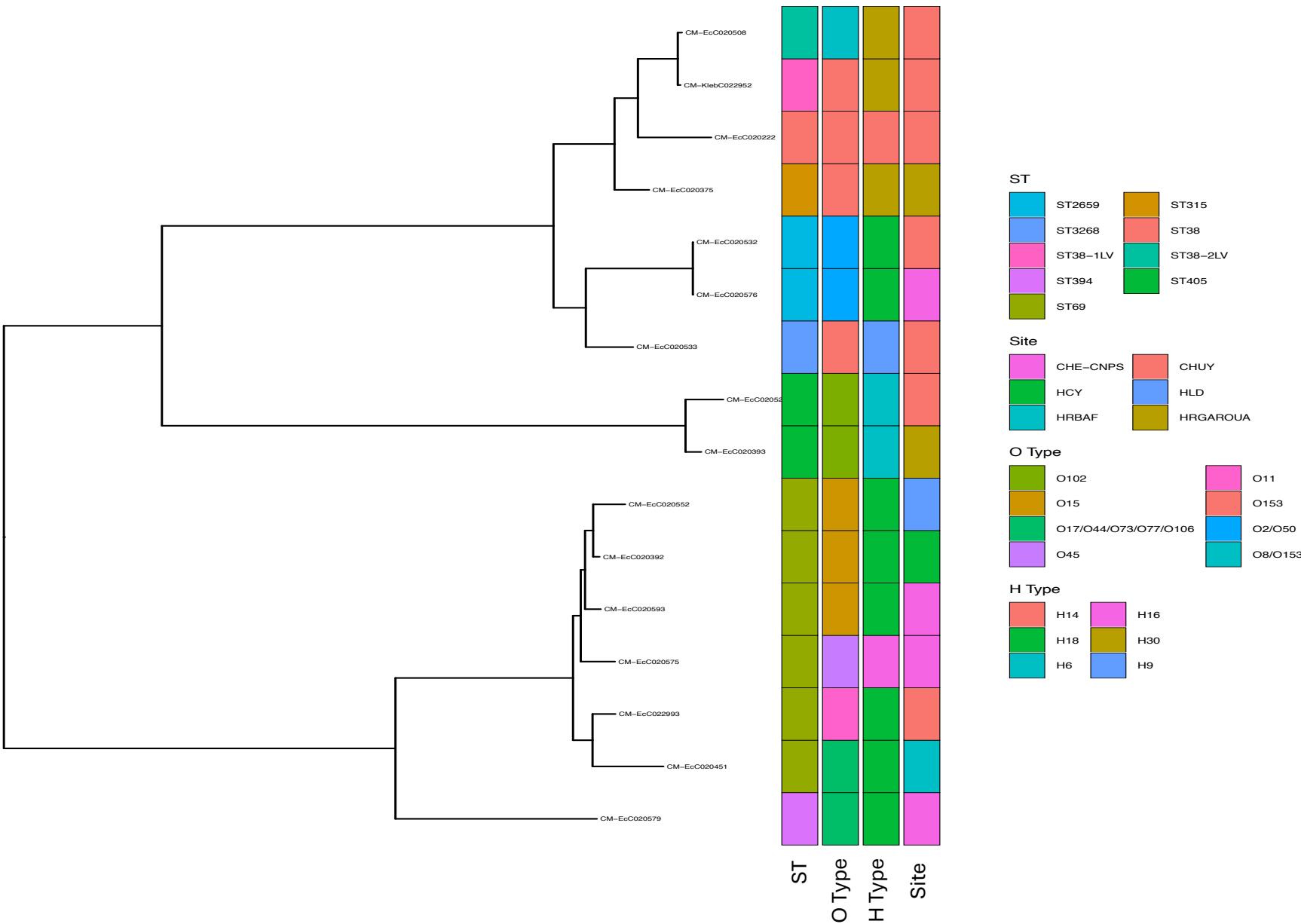
Results of phylogenetic analysis : Clade 2 (3/5)



Results of phylogenetic analysis : Clade 3 (4/5)



Results of phylogenetic analysis : Clade 4 (5/5)



Summary

- **287 sequences obtained after TheiaProk PHB analyses:** 159 KPSC, 126 *Escherichia coli/Shigella* and 2 *Klebsiella aerogenes*
- **Characterization of KPSC strains:**
 - Most prevalent STs: **ST101, ST39, ST147** and **ST307**
 - Most prevalent O types: **O1αβ,2β; O1αβ,2α** and **O2β**
 - Most prevalent K types: **K2** and **unknown (KL102)**
 - The strains were MDR, with a strong presence of **CTX-M-15, aac(3)-Ila, sul1** and **tet(A)** genes
 - The **IncFIB(K)** plasmid was the most frequently observed
 - All strains had a virulence score of 1, except for one strain with a virulence score of 3, characterized by:
 - **ST25-1LV, K-type K2, O-type O1αβ,2β** and carrying the **IncFIB(K)** and **IncHI1B** plasmids
 - Inter- and intra-laboratory transmissions were observed in some laboratories
- **Characterization of *Escherichia coli/Shigella* strains:**
 - Most prevalent STs: **ST131, ST1193, ST410** and **ST131-1LV**
 - Most prevalent O type: **O25**
 - Most prevalent H type: **H5**
 - The strains were MDR, with an important presence of **IncFIA** and **IncFIB(AP001918)** plasmids
 - Inter- and intra-laboratory transmissions were observed in some laboratories

A grid of 16 boxes for handwriting practice, containing various strokes and symbols.

< question ? >

```
\_ .--.  
 \ |o_o |\  
  \|:_/_|\  
   // \ \ \\\  
    (|     | )  
   /' \_ _/_` \\\  
  \___)=(_/_\
```