

# Implementation und Validation of Bioinformatics Pipelines for Routine Diagnostic based on Whole-Genome-Sequencing

Jörg Linde, Mostafa Abdel-Glil, Herbert Tomaso  
 Institute for Bacterial Infection and Zoonoses (IBIZ), Jena, Germany

High-throughput-sequencing (HTS) followed by bioinformatics data analysis is a powerful tool for genotyping and outbreak analysis. Though a number of protocols and bioinformatics tools exist, standardized lab protocols and easy-to-use open-source bioinformatics pipelines validated for guidelines of the International Standardization Organization (ISO) are needed to routinely perform HTS for genotyping and outbreak analysis. The IBIZ is responsible for German national reference laboratories of zoonotic bacteria. Here, we present our approach on implementation and validation of a HTS workflow according to the guidelines of ISO 17025 and show example results for genotyping of *Francisella tularensis* isolates.

## Task

HTS-based routine diagnostic for twelve German reference laboratories including:

- *Salmonella* spp.
- *Campylobacter* spp.
- *Brucella* spp.
- *Bacillus anthracis*
- *Francisella tularensis*

Perform genotyping based on

- MLST (classical and core-genome)
- canonical SNPs (CanSNPer)
- alignment-based SNPs (ParSNP)
- Kmers based SNPs(kSNP)

## Evaluation

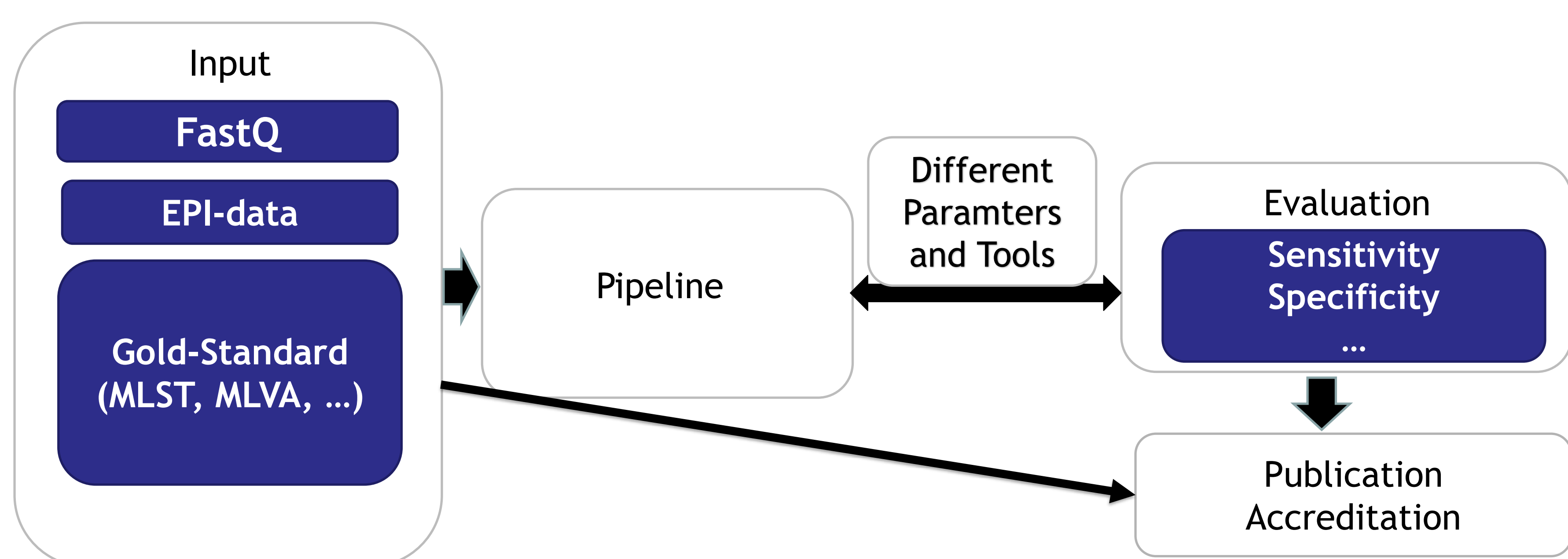


Figure 1 Evaluation scheme

## Implementation

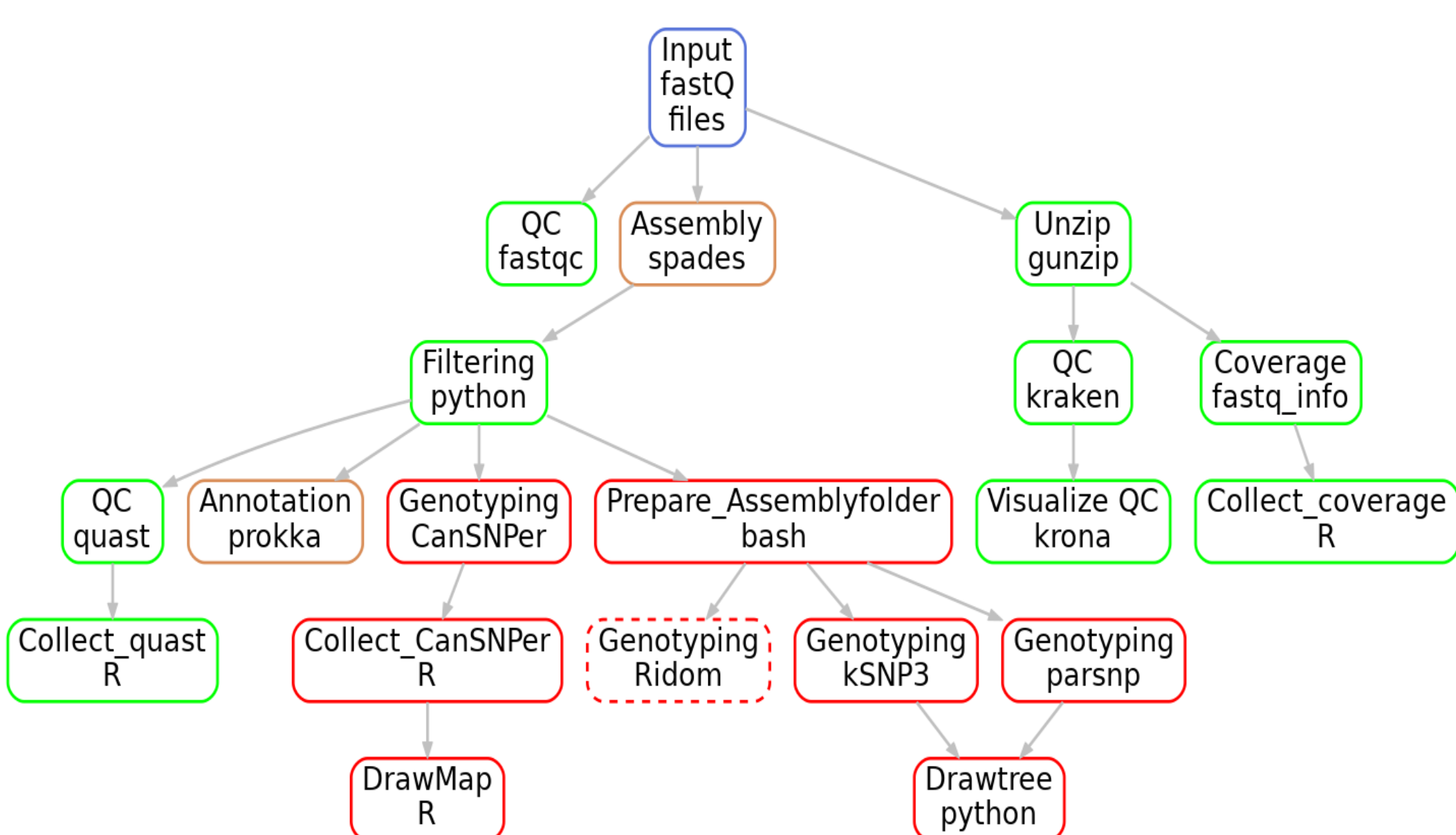


Figure 2: Pipeline uses **Snakemake** for scalability. Ridom is used externally for core-genome MLST.

```

rule parsnp:
  input:
    assemblies="results/assemblies",
    ref=config["Reference"]
  output:
    directory("results/parsnp")
  threads:
    8
  benchmark:
    "benchmarks/parsnp.txt"
  conda:
    "envs/parsnp.yaml"
  log:
    "log/parsnp.log"
  shell:
    "parsnp -r {input.ref} -d {input.assemblies} -p {threads} -o {output} && {log}"
  
```

Input data

Result

Number CPUs

Runtime

Automatic installation

Log

Command

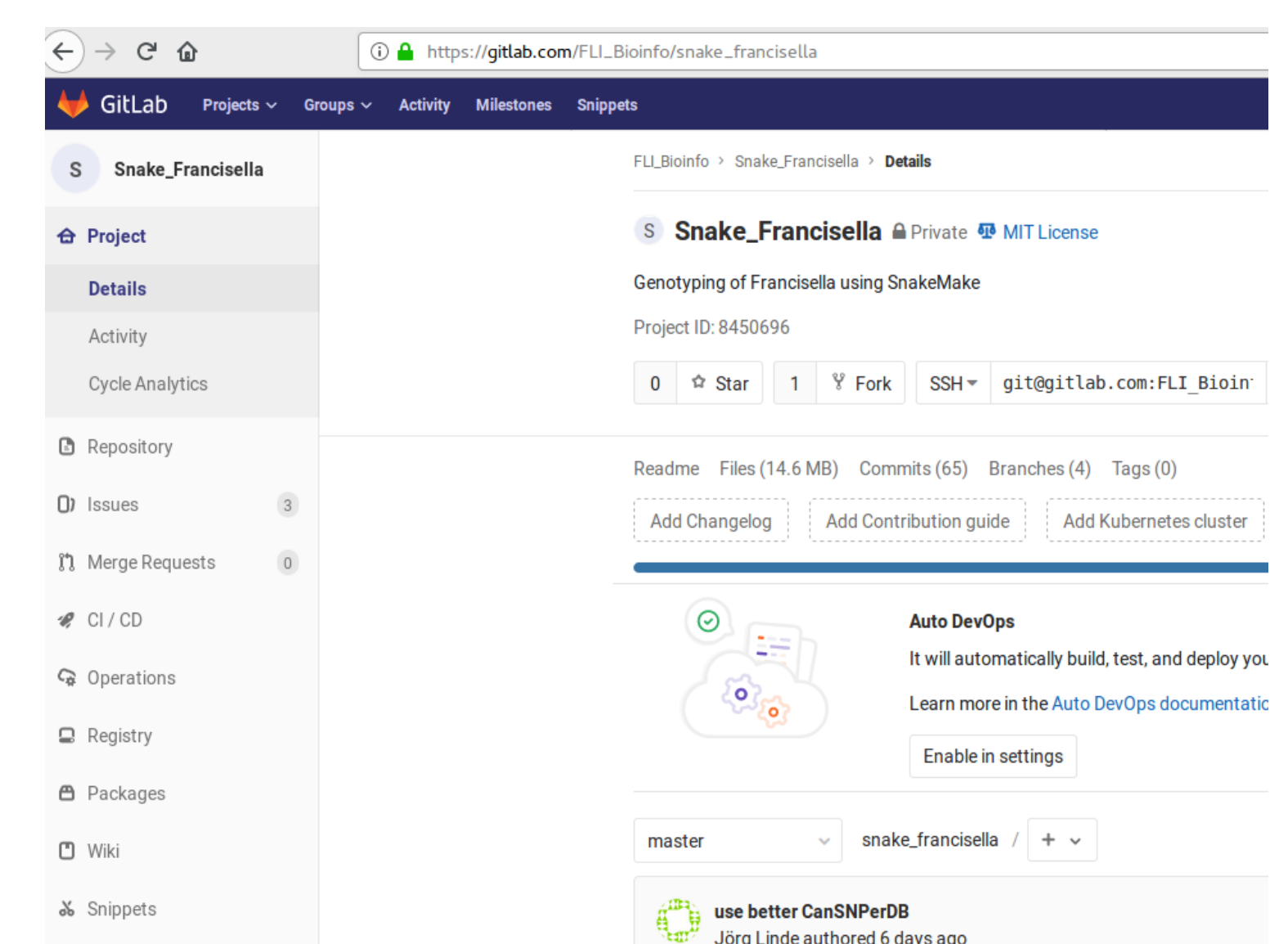


Figure 4: **Git** is used for version control and sharing.

## Example Results for *Francisella tularensis*

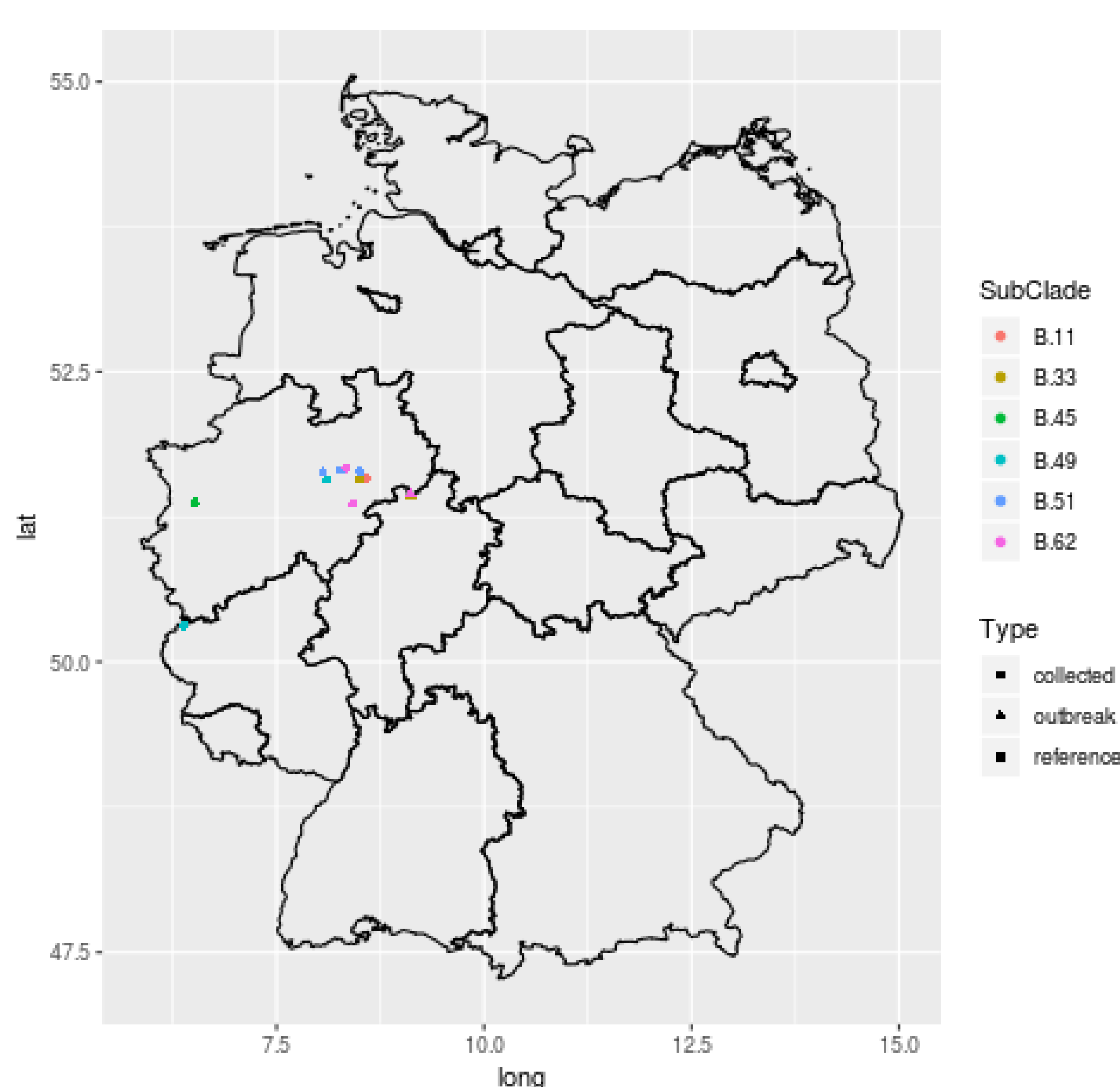


Figure 5: Isolates from North-Rhine-Westphalia (Busch *et al*, 2018) and their subclade from canonical SNPs.

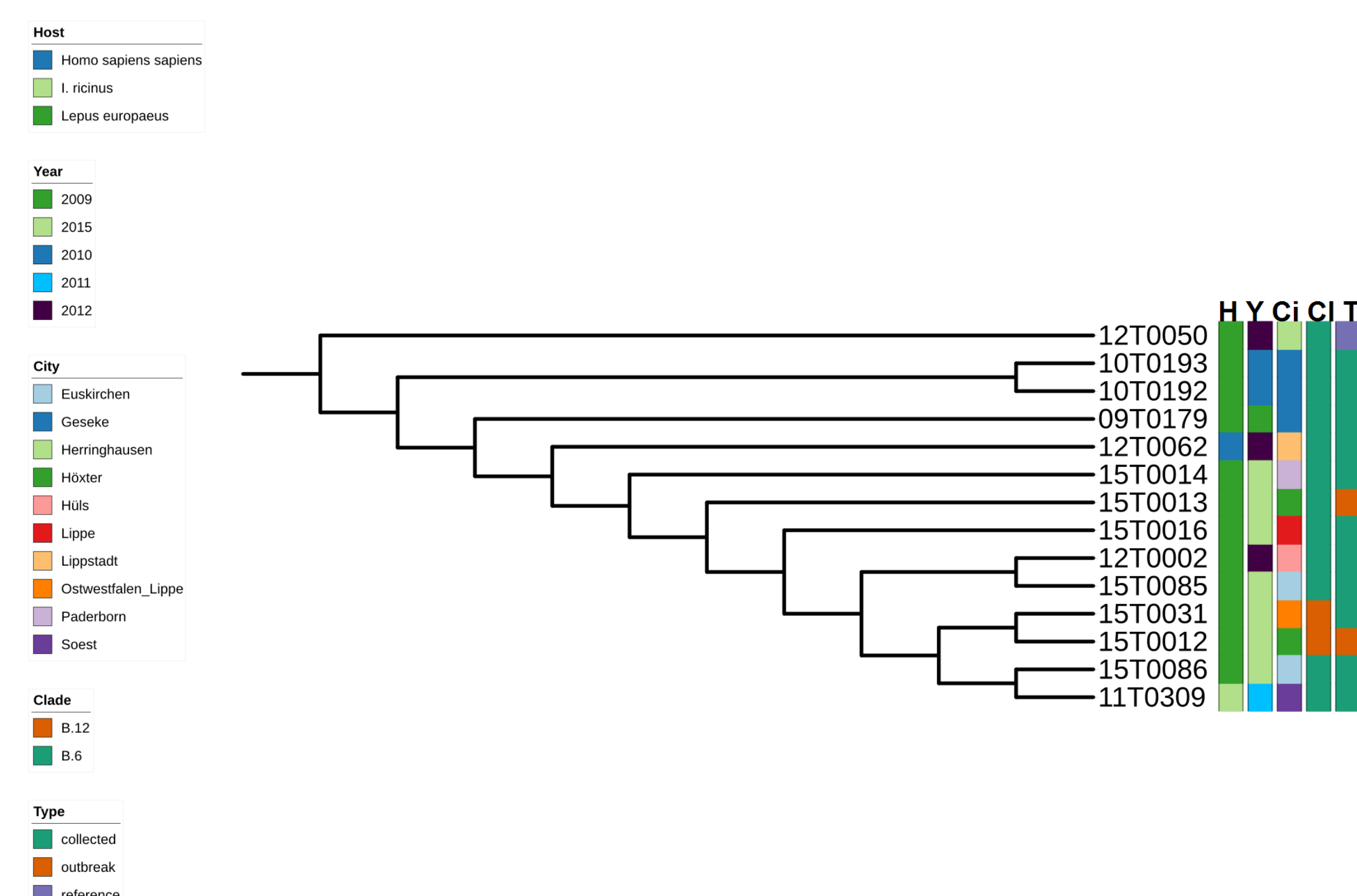


Figure 6: Example of SNP based tree (ParSNP) of isolates from North-Rhine-Westphalia (Busch *et al*, 2018).