

### 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 opensource 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.

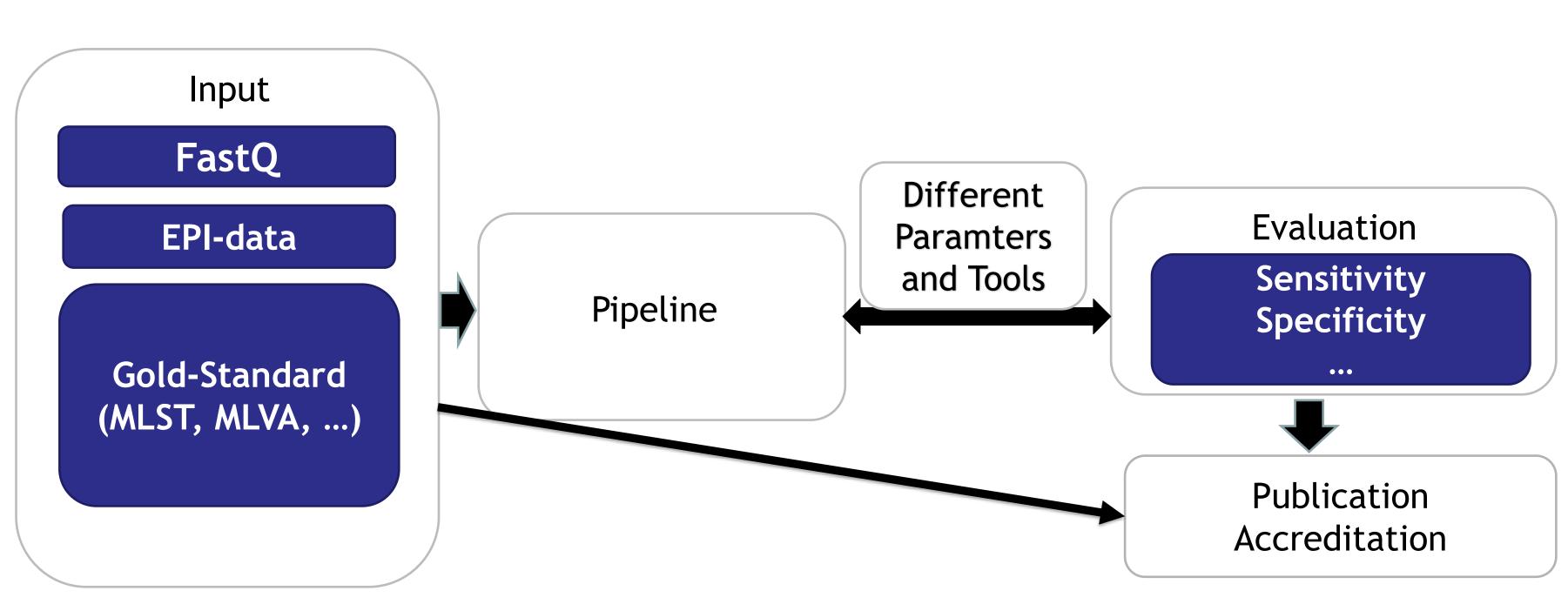
# Task

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

- Salmonella spp.
- *Campylobacter* spp. lacksquare
- Brucella spp.  $\bullet$
- 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** 

**Figugre 1** Evaluation scheme

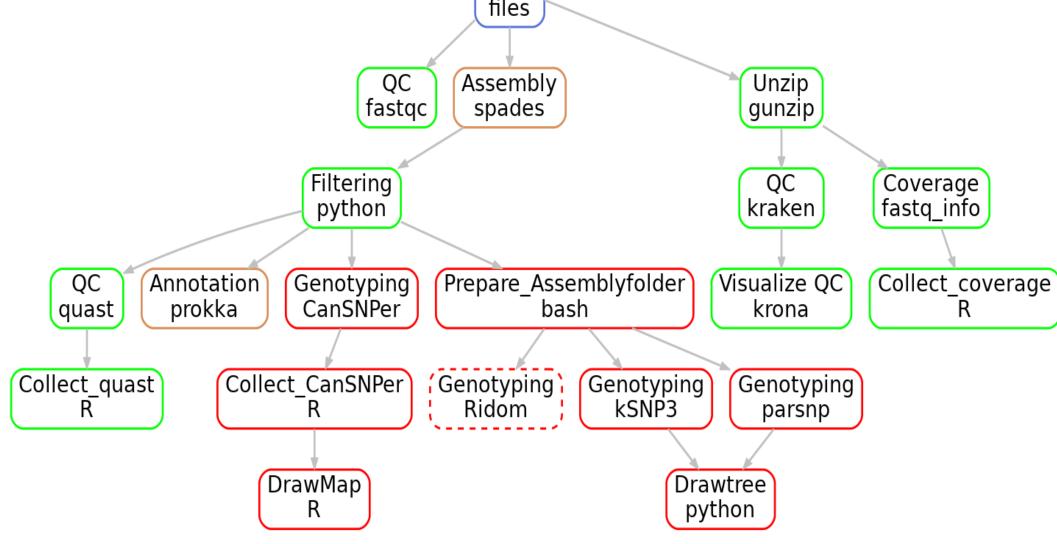
## Implementation



log:

 $(\leftarrow) \rightarrow$  C  $(\triangle)$ 

🕽 🔒 https://**gitlab.com**/FLI\_Bioinfo/snake\_francisella



input: assemblies="results/assemblies", Input data ref=config["Reference"] output: Result directory("results/parsnp") Number CPUs threads: Runtime benchmark: "benchmarks/parsnp.txt" Automatic installation conda: "envs/parsnp.yaml" Log "log/parsnp.log" Command shell: "parsnp -r {input.ref} -d {input.assemblies} -p {threads} -o {output} &> {log}

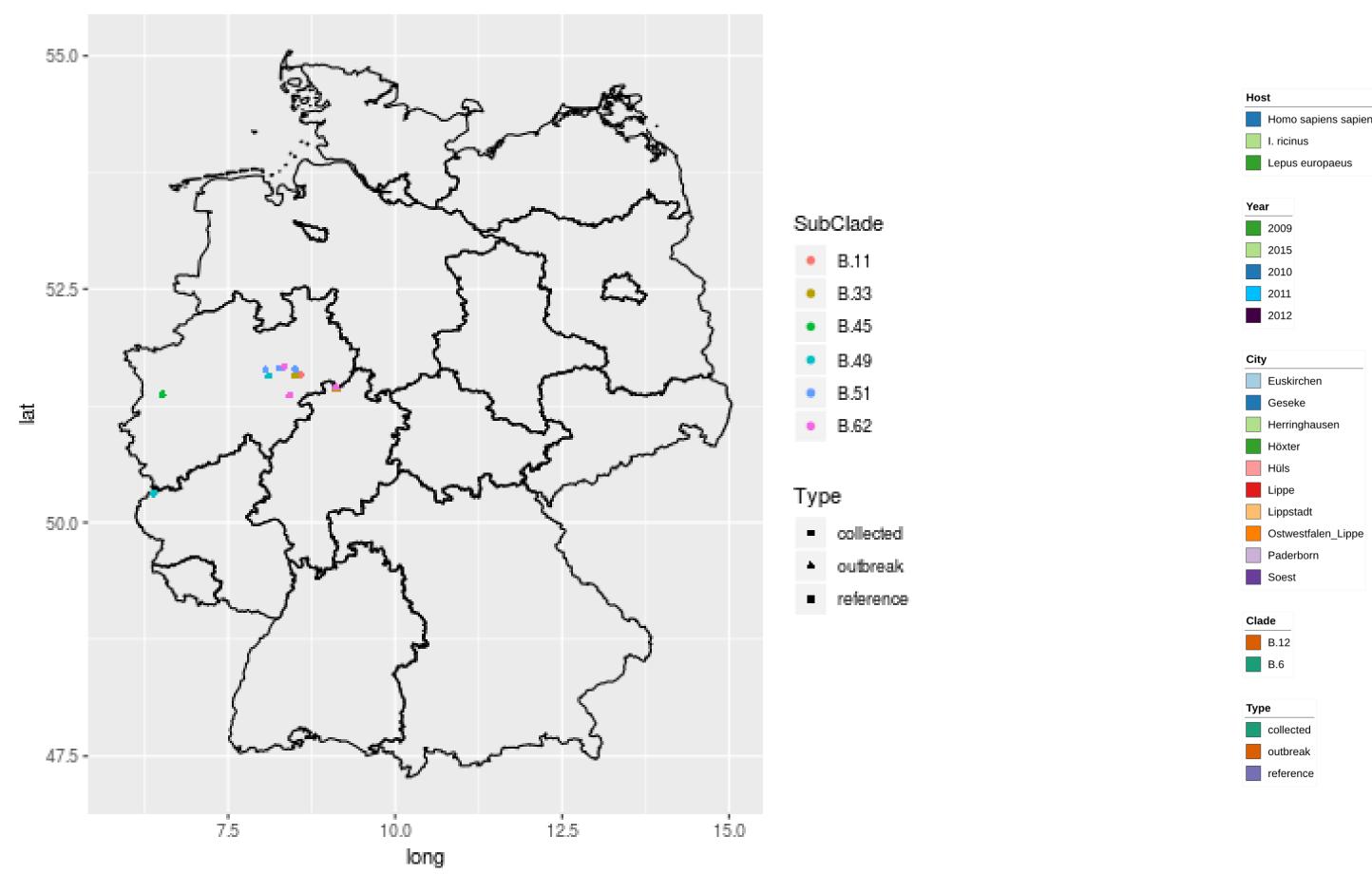
Figure 2: Pipeline uses Snakemake for scalability. Ridom is used externally for core-genome MLST.

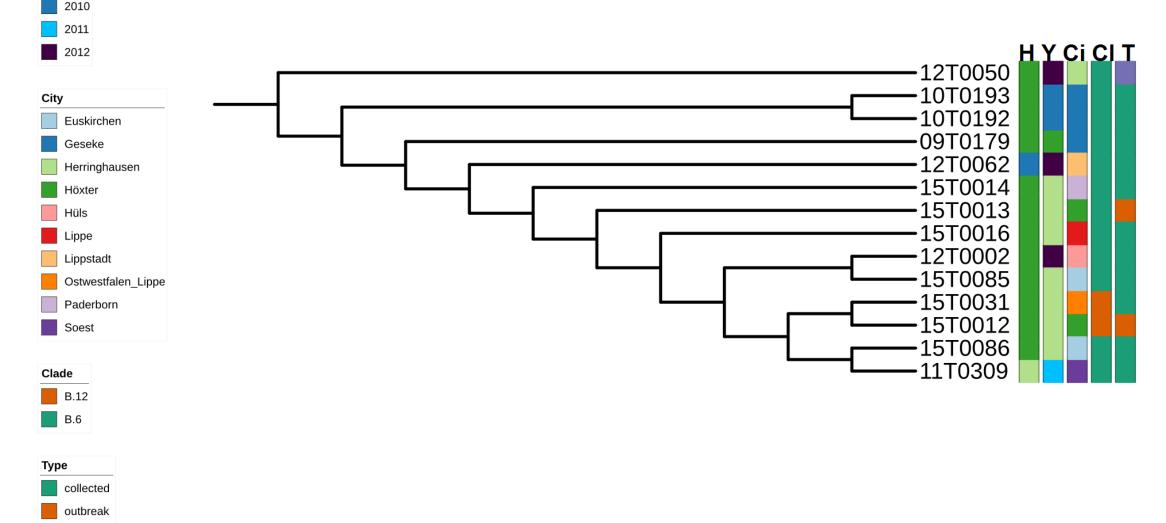
Figure 3: Snakemake uses (bio)conda for easy installation of software dependencies.

₩ GitLab Projects	~ Grou	ps 🗸 Activity	Milestones	Snippets
S Snake_Francisella				FLI_Bioinfo > Snake_Francisella > <b>Details</b>
🔂 Project				S Snake_Francisella A Private  MIT License
Details				Genotyping of Francisella using SnakeMake
Activity				Project ID: 8450696
Cycle Analytics				0 ☆ Star 1 ∛ Fork SSH ▼ git@gitlab.com:FLI_Bioin <sup>.</sup>
Repository				Readme Files (14.6 MB) Commits (65) Branches (4) Tags (0)
() Issues	3			Add Changelog Add Contribution guide Add Kubernetes cluster
រ៉ា Merge Requests	0			······································
🤗 CI/CD				O Auto DevOps
<table-cell> Operations</table-cell>			It will automatically build, test, and deploy you Learn more in the Auto DevOps documentation	
Registry			Enable in settings	
🖰 Packages				
🗂 Wiki				master v snake_francisella / + v
🕉 Snippets				use better CanSNPerDB Jörg Linde authored 6 days ago

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).