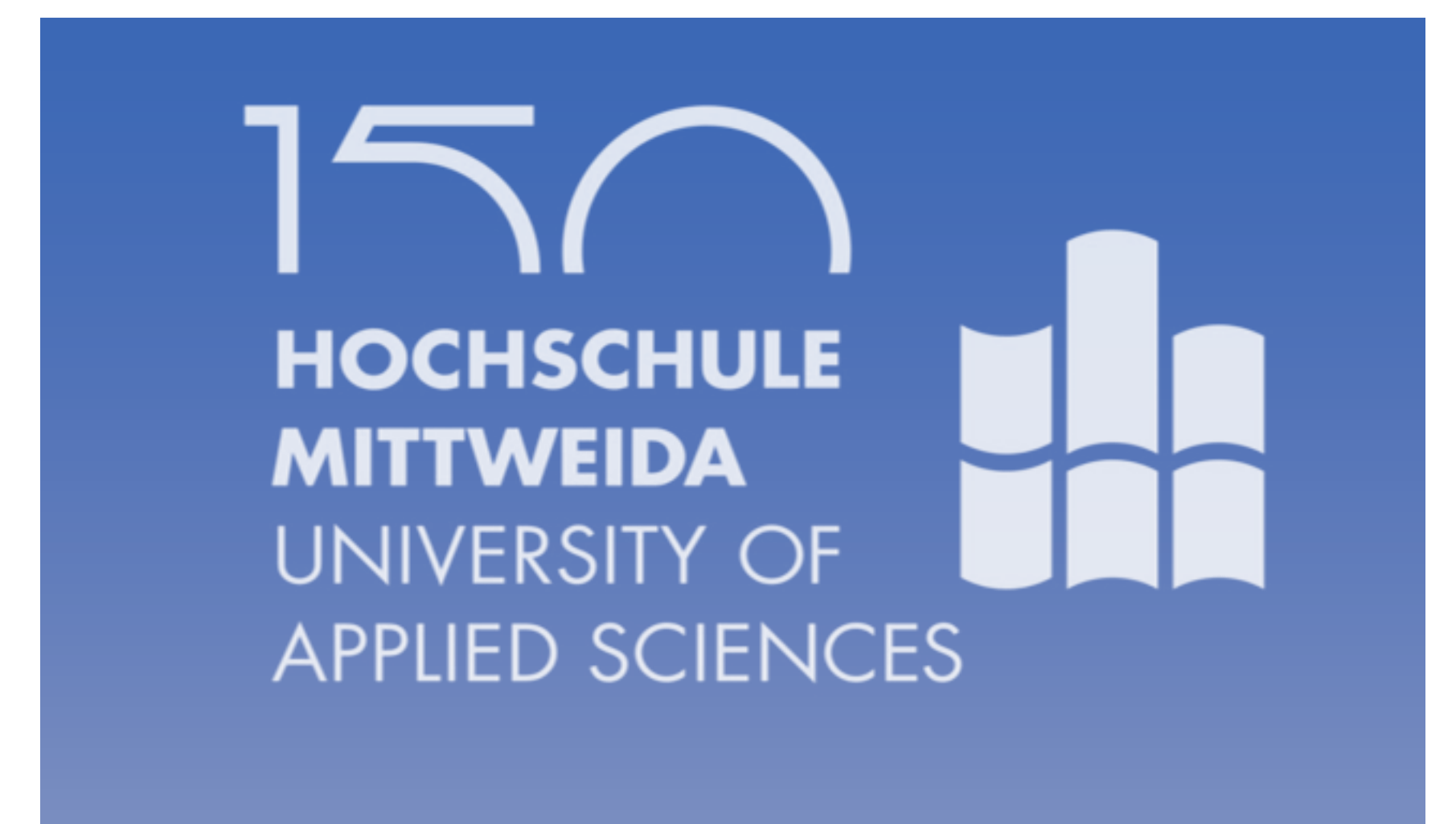


Forensische Mikrobiologie: Effiziente Klassifikation von Hasenpesterregern

Eric Zuchantke¹, Röbbbe Wünschiers¹, Herbert Tomaso²

¹ Biotechnologie/Chemie, Hochschule Mittweida

² Institut für bakterielle Infektion und Zoonosen, Friedrich-Loeffler-Institut, Jena



FRIEDRICH-LOEFFLER-INSTITUT

FLI

Bundesforschungsinstitut für Tiergesundheit
Federal Research Institute for Animal Health

Grundlage

Kanonische Einzelnukleotid-Polymorphismen (canSNPs) und Insertionen bzw. Deletionen (InDels) werden genutzt um die Verbreitung von *Francisella tularensis* zu analysieren und einzelne Stämme zu identifizieren. Durch Methoden wie der Multiplen Loci VNTR Analyse (MLVA) können diese Einzelnukleotid-Polymorphismen identifiziert werden. Für die Identifikation, ob sich ein canSNPs in einem mutierten oder nicht-mutierten Zustand befindet, werden mindestens 2 Primer- bzw. Marker-Paare benötigt. Da diese Laborarbeit in Anbetracht einer großen Anzahl an canSNPs zeitaufwendig und teuer ist, ist die Entwicklung und Nutzung eines computergestützten Ansatzes unausweichlich. Neben der Analyse mittels MLVA wird auch Next Generation Sequencing (NGS) zur Aufklärung des Genoms genutzt.

Durch die Nutzung vorhandener Sequenzinformationen von *F. tularensis* haben wir eine zeit- und kosteneffiziente Anwendung erzeugt, **CanFindIt**, mit der es möglich ist einzelne Stämme zu identifizieren und mit anderen Stämmen in Verbindung zu setzen.

Material und Methoden

CanFindIt ist ein Python geschriebener Webservice. Zum testen der Anwendung wurden 11 Proben von *Francisella tularensis* subsp. *holarctica* genutzt [1]. Für die Identifikation wurden sowohl die rohen Sequenzdaten aus den experimentell generierten FASTQ-Dateien genutzt, als auch die assemblierten Sequenzen im FASTA-Format. Abbildung 1 zeigt die Arbeitsweise des Programms.

Um die experimentellen Daten zu bestätigen wurden neben den canSNPs auch die InDels getestet. Um die gegebenen Proben zu identifizieren wurden spezifische Marker für canSNPs und InDels genutzt, welche in vorhergehenden Arbeiten [2,3,4,5] genutzt wurden.

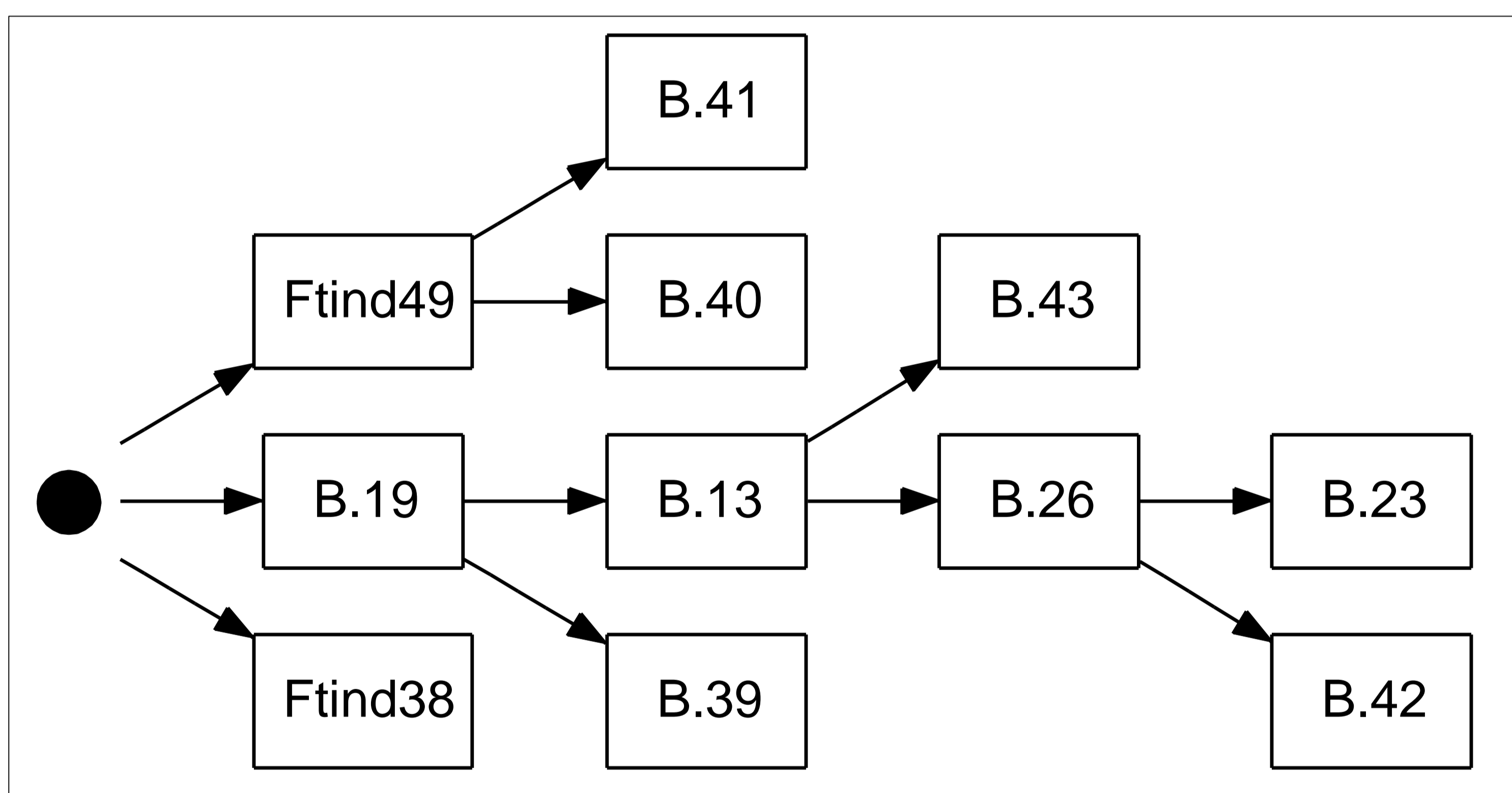
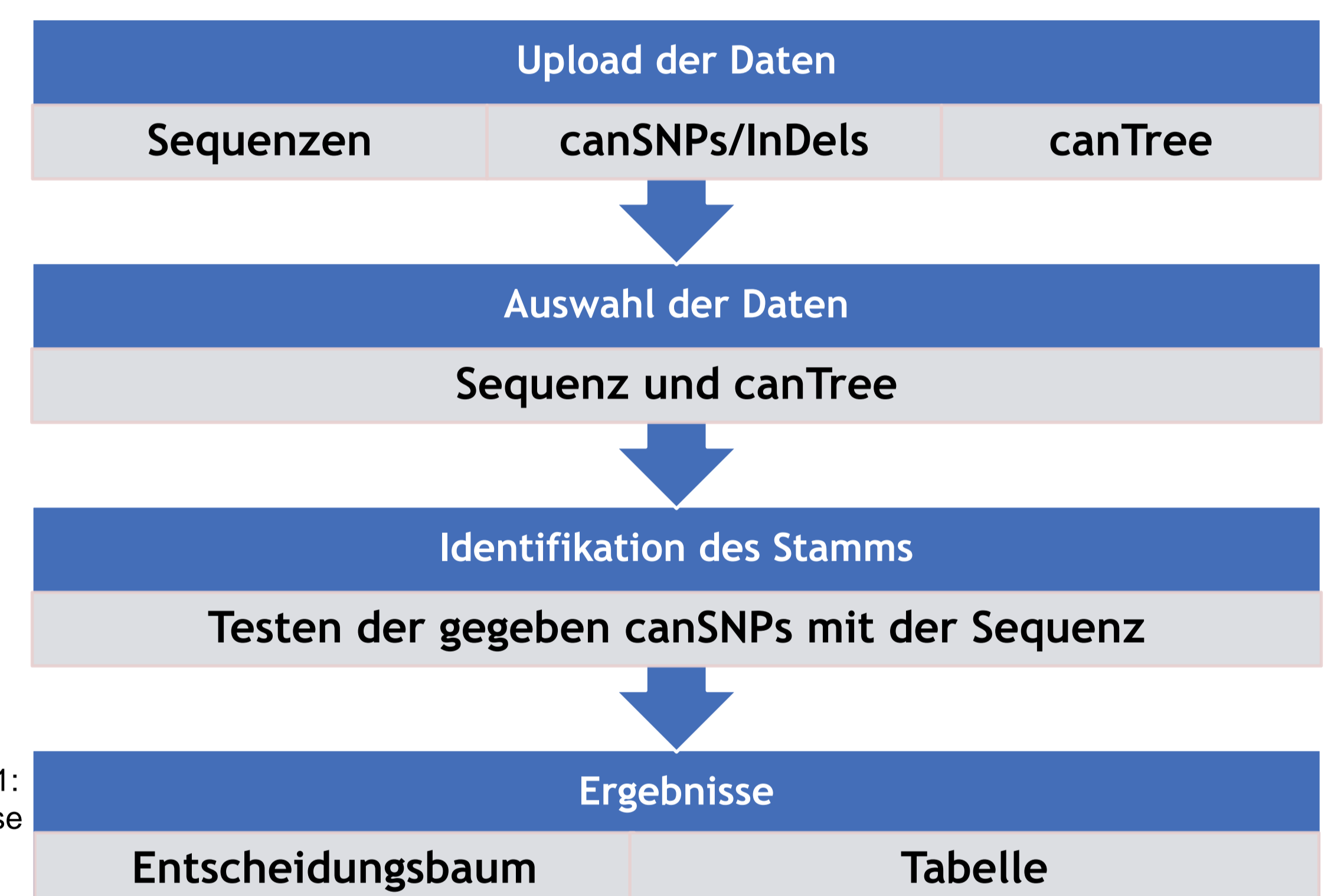


Abbildung 2:
CanTree, welches die Suchstrategie für die Stammidentifikation enthält. Aus Informationen von [1] erstellt.

Abbildung 1:
Arbeitsweise von CanFindIt



Für eine schneller Berechnung wird ein branch and bound Modell für die Identifikation genutzt. Das bedeutet, dass wie in richtigen Experimenten ein Entscheidungsbaum (canTree) erstellt wird. Dieser beinhaltet die Informationen über die CanSNPs und InDels und deren Zustände. Abbildung 2 zeigt einen solchen Entscheidungsbaum. Nur wenn ein canSNP oder InDel als mutiert angesehen wird, werden die darauffolgenden canSNPs oder InDels weiter untersucht.

Ergebnisse und Diskussion

Mittels **CanFindIt** war es möglich die experimentellen Ergebnisse zu rekonstruieren, wie Tabelle 1 zeigt. Die Webanwendung erlaubt eine schnelle Identifikation von gegebenen canSNPs und InDels, wobei nur die rohen Sequenzinformationen aus einer FASTQ-Datei genutzt werden können. Die Nutzung der assemblierten Sequenzen aus dem FASTA Format ist auch möglich, jedoch sind die Ergebnisse zum jetzigen Stand der Anwendungen weniger genau.

Um gute Ergebnisse durch **CanFindIt** zu erhalten, ist es nötig das die genutzte Sequenz von hoher Qualität ist und keine Kontamination enthält.

	Ftind38	Ftind49	B.19	B.23	B.13	B.40	B.41	B.42	B.43	B.26	B.39	
06T0001	A	A	D	A	D			D	A	D	A	B.I
08T0001	A	A	D	A	D			X	A	D	A	
09T0163	A	A	D	A	D			D	X	D	A	
09T0164	A	A	D	A	D			D	A	D	A	
08T0075	A	D	A			A	D					B.IV
09T0179	A	D	A			A	D					
10T0014	A	D	X		A	A	D				A	
12T0062	A	D	A			A	D					
13T0009	A	D	X		A	A	D				A	
13T0019	D	A	A		A						A	B.II
08T0013	D	A	X		A						A	

Tabelle 1: Vergleich der experimentellen Ergebnisse mit den Ergebnisse aus **CanFindIt**.

● korrekt identifizierte mutierte Position
● Position welche aufgrund des vorherigen Ergebnisses nicht getestet wurde
● Keine Übereinstimmung zwischen **canFindIt** und Experiment

Zusammenfassung und Ausblick

CanFindIt befindet sich in einem frühen Entwicklungsstadium, zeigt jedoch das die grundlegende Funktion, die Identifikation von canSNPs und auch InDels, mit hoher Genauigkeit funktioniert. In naher Zukunft werden neue Funktionen eingefügt, welche folgendes ermöglichen:

- Aufspürung neuer möglicher canSNPs für eine bessere Klassifikation innerhalb eines Stammes
- geographischer Vergleich von *Francisella tularensis* Proben
- automatische Assemblierung von rohen Sequenzinformationen
- automatische Erstellung phylogenetischer Bäume

Literatur

- [1] Karlsson, E., Svensson, K., et al. 2012. The phylogeographic pattern of *Francisella tularensis* subsp. *holarctica* in Sweden indicates a Scandinavian origin of Eurosiberian tularaemia. *Environ. Microbiol.* DOI:10.1111/1462-2920.12052
- [2] Chanturia, G., Birdsell D. N., et al. 2011. Phylogeography of *Francisella tularensis* subspecies *holarctica* from the country of Georgia. *BMC Microbiol* 11:139.
- [3] Larsson, P., K. Svensson, L., et al. 2007. Canonical insertion-deletion markers for rapid DNA typing of *Francisella tularensis*. *Emerg Infect Dis* 13:1725-1732.
- [4] Svensson, K., M. Granberg, L., et al. 2009. A real-time PCR array for hierarchical identification of *Francisella* isolates. *PLoS One* 4:e8360.
- [5] Vogler, A. J., D. Birdsell, L. B., et al. 2009. Phylogeography of *Francisella tularensis*: global expansion of a highly fit clone. *J Bacteriol* 191:2474-2484.