

Aus dem Forschungszentrum Borstel
Forschungsgruppe Molekulare und Experimentelle Mykobakteriologie
Programmbereich Infektion
Direktor: Prof. Dr. Ulrich Schaible

Nachweis von Antibiotikaresistenzen mittels Genomanalyse von Tuberkulosebakterien

Detektion heterogener Populationen und deren Bedeutung für die
Resistenzvorhersage

Inauguraldissertation
zur
Erlangung der Doktorwürde
der Universität zu Lübeck



UNIVERSITÄT ZU LÜBECK

vorgelegt von
Viola Schleusener
aus Eutin

Lübeck 2017

1. Berichterstatter: Prof. Dr. Stefan Niemann

2. Berichterstatter/Berichterstatterin: Prof. Dr. Werner Solbach

Tag der mündlichen Prüfung: 17.08.2017

Zum Druck genehmigt. Lübeck, den 20.02.2018

Zusammenfassung

Die Tuberkulose (TB) zählt mit circa 10 Millionen Neuerkrankungen und 1,4 Millionen Todesfällen in 2015 zu den bedeutendsten Infektionskrankheiten weltweit. Besonders die steigenden Zahlen resistenter und multiresistenter (MDR) Fälle erschweren die Behandlung und stellen ein zunehmendes Problem für die Gesundheitssysteme dar. Eine frühzeitige Erkennung vorhandener Antibiotikaresistenzen ist der Schlüssel für eine adäquate Therapie. Jedoch kann die traditionelle phänotypische Resistenztestung auf Grund des langsamen Wachstums des TB-Erregers bis zu acht Wochen dauern. Genomsequenzierungstechniken der nächsten Generation (NGS, engl.: next-generation sequencing; NGS) bieten neue Möglichkeiten resistenzvermittelnde Mutationen zuverlässig zu bestimmen und damit ein für den Patienten individualisiertes, molekularbiologisches Resistenzprofil zu erstellen.

Ein Problem bei der Anwendung der NGS-basierten Diagnostik ist die komplexe Prozessierung, Aufbereitung und Interpretation der Daten. Zusätzlich erschweren heterogene Populationen, beispielhaft ausgelöst durch Mischinfektionen zweier unterschiedlicher TB-Stämme oder durch die Mikroevolution des Pathogens im Wirt und die damit zum Teil einhergehenden Heteroresistenzen, die Interpretation der Daten.

Der NGS-basierte Nachweis heterogener Populationen in klinischen TB-Proben ist bis jetzt noch nicht ausreichend untersucht und beschrieben worden.

Ein Ansatz für die breitere Anwendung NGS-basierter Diagnostik ist die Anwendung automatisierter Analyseprogramme. Hier stehen, neben der in dieser Arbeit mit entwickelten Webseite PhyResSE, weitere frei-verfügbare Webseiten oder Analyseprogramme zur Verfügung. Eine vergleichende Evaluierung der Resistenzvorhersage wurde allerdings noch nicht durchgeführt. Um diese Fragestellung zu adressieren, wurden in dieser Arbeit die für die TB-Resistenzdiagnostik entwickelten, automatisierten NGS-Analyseprogramme CASTB, KvarQ, Mykrobe Predictor TB, TBProfiler und PhyResSE systematisch evaluiert. Zudem wurden in dieser Arbeit neuartige Verfahren implementiert, die zum einen heterogene Populationen (Mischinfektionen und niederfrequente Resistenzmutationen) nachweisen und zum anderen die detektierten, niederfrequenten Resistenzmutationen mittels eines statistischen Verfahrens bewerten, um sie für eine valide Diagnostik zugänglich zu machen.

Der Vergleich der automatisierten NGS-Analyseprogramme zeigte, dass die Sensitivität der Resistenzvorhersage zwischen den fünf untersuchten Programmen stark variiert; so lagen zum Beispiel die Werte für den Nachweis von Isoniazid Resistenz zwischen 100 % für PhyResSE und 86 % für CASTB und KvarQ. Die Werte für den Ausschluss von Isoniazid Resistenz lagen zwischen 100 % für CASTB, KvarQ und Mykrobe Predictor TB und 85 % für TBProfiler, das sieben Isoniazid sensible TB-Stämme fehlinterpretierte. Die größte Diskrepanz zwischen den untersuchten Programmen konnte für Streptomycin mit Sensitivitäten zwischen 30 % (CASTB) und 86 % (PhyResSE) beobachtet werden.

Da mit keinem der untersuchten, automatisierten NGS-Analyseprogramme eine zuverlässige Detektion heterogener Populationen möglich war, wurde in dieser Arbeit ein Verfahren implementiert, das basierend auf dem Clusteralgorithmus DBSCAN, Schwerpunkte in der Häufigkeitsverteilung der Allelfrequenzen detektiert. Die Sensitivität und Spezifität des entwickelten Mischungsdetektors wurde anhand von 1200 Genomdatensätzen ermittelt und ergab für Mischinfektionen, bei denen die schwächere Population mindestens einen Anteil von 5% ausmachte, eine Sensitivität von 100%, sowie eine Spezifität von 97%.

Des Weiteren wurde ein Verfahren entwickelt, das niederfrequente Resistenzmutationen detektiert und mit Hilfe des Binomialtests bewertet. Es konnten bei einer Abdeckung von 400 NGS-Sequenzfragmenten Mutationen bis zu einer Frequenz von 1% valide detektiert werden.

Beide Verfahren wurden sowohl mit artifiziellen NGS-Datensätzen, als auch mit klinischen NGS-basierten Daten überprüft.

Anhand serieller NGS-Daten von 13 Patienten konnte gezeigt werden, dass Reininfektionen und Mischinfektionen mit unterschiedlichen TB-Stämme eine große Bedeutung bei Therapieversagen haben. So wurde beispielhaft bei 10 der 13 Patienten eine Re- oder Mischinfektion mit Hilfe des hier implementierten Verfahren nachgewiesen, welche mit klassischen Methoden (Gold Standards) nicht detektierbar waren.

Zusammengefasst konnte in dieser interdisziplinäre Arbeit die Etablierung neuartiger bioinformatischer Methoden mit der Anwendung NGS-basierter TB-Diagnostik verbunden werden. Eine verbesserte und optimierte Interpretation von NGS-Datensätzen mit Hilfe der hier implementierten Algorithmen leistet einen entscheidenden Beitrag auf dem Weg zur personalisierten Therapie von TB-Patienten im klinischen Alltag.

Inhaltsverzeichnis

Abbildungsverzeichnis	xii
Tabellenverzeichnis	xiv
Liste der Quelltexte	xvii
Abkürzungsverzeichnis	xviii
I Einführung	1
1 Tuberkulose	2
1.1 Geschichte der Tuberkulose	2
1.2 Epidemiologie der Tuberkulose	4
1.3 Die Tuberkuloseinfektion	8
1.4 <i>M. tuberculosis</i> -Komplex	8
1.5 Populationsstruktur des MTBK	9
1.6 Diagnose und Therapie der Tuberkulose	13
1.7 Heteroresistente und Mischpopulationen	17
1.8 Genomsequenzierung von MTBK-Stämmen	20
1.9 Zielsetzung	25
2 Grundlagen	27
2.1 Illumina-Sequenzierung	27
2.2 Ausgabeformate	27
2.2.1 FastQ-Format	29
2.2.2 Sequenz-Ausrichtungs-Format (SAM)	31
2.2.3 Varianten-Format (VCF)	33
2.2.4 Komma-getrennte Werte-Format (CSV)	33
2.2.5 JavaScript-Objekt-Format (JSON)	34
2.3 Analyse von NGS-Daten	34

2.4	Verwendete Datensätze	36
2.5	Verwendete Software und Programmiersprachen	40
2.5.1	Statistikprogramm R	40
2.5.2	Bam-readcount Algorithmus	40
2.5.3	Programmiersprache Perl	41
II Ergebnisteil		42
3	Automatisierte NGS-Analyseprogramme	43
3.1	CASTB (Version 1.1)	43
3.2	KvarQ (Version 0.12.2)	45
3.3	Mykrobe Predictor TB (Version 0.1.3)	46
3.4	PhyResSE (Version 1.0)	46
3.5	TBProfiler	47
4	Evaluierung automatischer NGS-Analyseprogramme mit Hilfe eines Validierungsdatensatzes	49
4.1	Allgemeiner Vergleich	49
4.2	Evaluierung der phylogenetischen Klassifizierung	54
4.3	Evaluierung der Resistenzvorhersage	55
4.4	Detektion von Mischinfektionen	61
4.5	Detektion von Heteroresistenzen	63
5	Heterogene Populationen	69
5.1	Detektor für Mischinfektionen	69
5.1.1	DBSCAN-Algorithmus	72
5.1.2	Parametereinstellung des DBSCAN-Algorithmus	73
5.1.3	Validierung des Mischungsdetektors	80
5.1.4	Untersuchung verschiedener Populationen	84
5.2	Heteroresistenzen	94
5.2.1	Binomialansatz	94
5.2.2	Binomialverteilung	95
5.2.3	Abdeckungsuntersuchung	97
5.2.4	Binomialansatz für artifizielle, heteroresistente NGS-Proben	99
5.2.5	Auswertung serieller Proben von TB-Patienten	102

III Diskussion und Ausblick	113
6 Diskussion	114
6.1 Vergleich automatischer Analyseprogramme für unprozessierte NGS-Daten	114
6.2 Detektion heterogener Populationen	122
6.2.1 Mischungsdetektor	122
6.2.2 Binomialtest	126
6.3 Fazit und Ausblick	131
IV Anhang	133
A ASCII-Kodierung	134
B Datensätze	135
B.1 Datensatz aus Sierra Leone	135
B.2 Datensatz aus Russland	136
B.3 Datensatz aus Kenia	138
B.4 Datensatz aus Swasiland	139
B.5 Datensatz aus Turkmenistan	139
B.6 Datensatz aus Usbekistan	140
C Beispiel für eine Positionsliste des Bam-readcount-Algorithmus	141
D Mutationslisten	142
D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm	142
E Berechnung von Sensitivität und Spezifität	187
E.1 Vier-Felder Tafeln zur Berechnung der Sensitivitäten und Spezifitäten basierend auf den Sanger-Ergebnissen)	187
E.1.1 Isoniazid Resistenzbestimmung	187
E.1.2 Rifampicin Resistenzbestimmung	190
E.1.3 Ethambutol Resistenzbestimmung	194
E.1.4 Streptomycin Resistenzbestimmung	197
E.1.5 Pyrazinamid Resistenzbestimmung	200

E.2	Vier-Felder Tafeln zur Berechnung der Sensitivitäten und Spezifitäten basierend auf der phänotypischen Resistenztestung)	203
E.2.1	Isoniazid Resistenzbestimmung	203
E.2.2	Rifampicin Resistenzbestimmung	207
E.2.3	Ethambutol Resistenzbestimmung	210
E.2.4	Streptomycin Resistenzbestimmung	213
E.2.5	Pyrazinamid Resistenzbestimmung	217
E.3	Sensitivität und Spezifität der einzelnen Prgogramme im Vergleich zur phänotypischen DST	219
E.4	Vorab veröffentlichte Sensitivitäten und Spezifitäten der automatisierten Programme	220
F	Untersuchung der seriellen NGS-Daten aus Usbekistan	223
F.1	Patient A	223
F.2	Patient B	226
F.3	Patient C	228
F.4	Patient D	229
F.5	Patient E	230
F.6	Patient F	230
F.7	Patient G	234
F.8	Patient H	235
F.9	Patient I	238
F.10	Patient J	240
F.11	Patient K	241
F.12	Patient L	244
F.13	Patient M	246
G	Nachsequenzierte Proben mit nicht eindeutigen MIRU-VNTR Typisierungsmustern	250
	Literaturverzeichnis	252
	Lebenslauf	288

Abbildungsverzeichnis

1.1	Geschätzte TB-Inzidenzraten für das Jahr 2015	5
1.2	Anteil neuer MDR/RR-TB-Fälle an allen TB-Fällen pro Region im Jahr 2015	7
1.3	Neighbour-Joining Baum zur Darstellung der globalen phylogene- tischen Struktur des MTBK	10
1.4	Detektion von Mischinfektionen mit den standardisierten Typisie- rungsmethoden.	19
2.1	Schematische Darstellung der Illumina DNS-Bibliothekenerstellung	28
2.2	Zusammenhang der Basenqualitätswerte mit der Fehlerwahrschein- lichkeit	32
2.3	Abfolge der Analyseschritte von PhyResSE mit zugehörigen Pro- grammen und resultierenden Formaten.	37
4.1	Position 761161 einer heteroresistenten NGS-Probe	64
4.2	Abdeckungsgrafik von KvarQ für die Kernregion des <i>rpoB</i> -Gens .	65
4.3	Sanger-Ergebnis der niederfrequenten EMB-Resistenzmutation . .	67
4.4	Ausrichtung der NGS-Sequenzfragmente an der Position 4247431 mit unterschiedlichen Analyseabfolgen	68
5.1	Streudiagramme der Allelfrequenzen der detektierten SNPs für die NGS-Proben Mix50, Mix70, Mix90 und eine Kontrolle	71
5.2	Schematische Darstellung des DBSCAN-Algorithmus	73
5.3	Sortierte k-Abstandsgrafiken für die artifiziellen Mischungen Mix50, Mix70, Mix90 und eine Kontrolle	75
5.4	Ergebnis der Clusteranalyse des DBSCAN-Algorithmus für die NGS- Proben Mix50, Mix70, Mix90 und eine Kontrolle.	76
5.5	Perl-Skript <code>Batch_mixDetector_coregenome.pl</code>	78
5.6	Ergebnis der Clusteranalyse des DBSCAN-Algorithmus für NextSeq- Daten mit $\epsilon = 0.05$ und $MinPts = 100$	79

5.7	Ergebnis der Clusteranalyse des DBSCAN-Algorithmus für NextSeq-Daten mit $\epsilon = 0.06$ und $MinPts = 55$	81
5.8	Inzidenzraten und Lage der untersuchten Regionen.	84
5.9	Darstellung der minimal detektierbaren Frequenz einer mutierten Base in Abhängigkeit der vorliegenden Abdeckung.	98
5.10	Schematische Darstellung des Perl-Skripts zur Bewertung niederfrequenter Resistenzmutationen.	100
5.11	Entwicklung des Resistenzprofils von Patient B anhand von NGS-Daten	106
5.12	Entwicklung des Resistenzprofils von Patient D anhand von NGS-Daten	108
5.13	Venn-Diagramm der VCF-Dateien von Patient D	109
5.14	Entwicklung des Resistenzprofils von Patient I anhand von NGS-Daten	110
5.15	Venn-Diagramm der VCF-Dateien von Patient I	110
5.16	Entwicklung des Resistenzprofils von Patient C anhand von NGS-Daten	111
5.17	Venn-Diagramm der VCF-Dateien von Patient C	112
B.1	Entwicklung der TB-Inzidenzrate in Sierra Leone	136
B.2	Genotypverteilung der Studienpopulation aus Russland.	138
F.1	Entwicklung des Resistenzprofils von Patient A anhand von NGS-Daten	224
F.2	Venn-Diagramm der VCF-Dateien der drei NGS-Proben für Patient A	225
F.3	Venn-Diagramm der VCF-Dateien der ersten drei NGS-Proben von Patient B	227
F.4	Venn-Diagramm der VCF-Dateien von Patient E	231
F.5	Entwicklung des Resistenzprofils von Patient E anhand von NGS-Daten	231
F.6	Entwicklung des Resistenzprofils von Patient F anhand von NGS-Daten	232
F.7	Venn-Diagramm der VCF-Dateien von Patient F	233
F.8	Entwicklung des Resistenzprofils von Patient G anhand von NGS-Daten	234

F.9 Venn-Diagramm der VCF-Dateien von Patient G	236
F.10 Entwicklung des Resistenzprofils von Patient H anhand von NGS-Daten	237
F.11 Venn-Diagramm der VCF-Dateien von Patient H	239
F.12 Entwicklung des Resistenzprofils von Patient J anhand von NGS-Daten	241
F.13 Venn-Diagramm der VCF-Dateien von Patient J	242
F.14 Entwicklung des Resistenzprofils von Patient K anhand von NGS-Daten	244
F.15 Venn-Diagramm der VCF-Dateien von Patient K	245
F.16 Entwicklung des Resistenzprofils von Patient L anhand von NGS-Daten	246
F.17 Venn-Diagramm der VCF-Dateien von Patient L	247
F.19 Venn-Diagramm der VCF-Dateien von Patient M	249
F.18 Entwicklung des Resistenzprofils von Patient M anhand von NGS-Daten	249

Tabellenverzeichnis

1.1	Die wichtigsten Antituberkulotika und ihre Wirkmechanismen . . .	16
1.2	Sequenzierungsmethoden im Vergleich	24
2.1	Qualitätswerte (Q) mit zugehörigen Fehlerwahrscheinlichkeiten (e) und korrespondierenden ASCII-Zeichen	30
2.2	Übersicht der verwendeten Datensätze	38
2.3	Zusammenfassung der in dieser Arbeit verwendeten, artifizielle DNS- Proben	39
4.1	Antibiotika für die eine Resistenzvorhersage erstellt wird	51
4.2	Funktionalität der fünf Programme	52
4.3	Phylogenetische Klassifizierung der MTBK-Stämme des Sierra Leo- ne Datensatzes	58
4.4	Detektierte Mutationen der evaluierten Programme	59
4.5	Sensitivität und Spezifität der Resistenzvorhersage verglichen mit der Sanger-Sequenzierung für die einzelnen Antibiotika im Test . .	60
4.6	Phylogenetische Klassifizierung der artifiziellen Mischinfektionen Mix50, Mix70 und Mix90	61
4.7	Ergebnisse der automatisierten NGS-Analyseprogramme für die ar- tifiziellen, heteroresistenten NGS-Proben.	66
5.1	Vierfeldertafel des Mischungsdetektors für die Russland-Studie (ENA Zugriffsnummer: ERP000192).	83
5.2	Vorkommen der verschiedenen phylogenetischen Linien in den he- terogenen NGS-Proben der Russland-Population.	85
5.3	Mischinfektionen unterschiedlicher Populationen	85
5.4	Mischdetektion in einer Population aus Kenia.	87
5.5	Stämme mit einer Mischinfektion aus einer Population aus Swasiland.	89
5.6	Mischdetektion in einer Population aus Turkmenistan.	91
5.7	Mischdetektion in einer Population aus Usbekistan (Nukus).	93

5.8	Ergebnis des Binomialtests für die Mischungen mit dem SR4k-Selektionsklon	101
5.9	Ergebnis des Binomialtests für die Mischungen mit dem SR1a-Selektionsklon	101
5.10	Phänotypische DST-Resistenzprofile von 13 Patienten aus Usbekistan	103
5.11	Analyse der NGS-Daten von 13 Patienten aus Usbekistan	105
A.1	ASCII-Zeichen für die Werte 33 bis 126	134
B.1	Resistenzprofil der Studienpopulation aus Sierra Leone	137
E.1	Sensitivität und Spezifität der Resistenzvorhersage verglichen mit den phänotypischen DST-Ergebnissen für die einzelnen Antibiotika im Test	221
E.2	Sensitivitäten und Spezifitäten der automatisierten Programme	222
F.1	Entwicklung des phänotypischen Resistenzprofils der Probe von Patient A	223
F.2	Entwicklung des phänotypischen Resistenzprofils der Probe von Patient B	226
F.3	Entwicklung des phänotypischen Resistenzprofils der Probe von Patient C	228
F.4	Entwicklung des phänotypischen Resistenzprofils der Probe von Patient D	229
F.5	Entwicklung des phänotypischen Resistenzprofils der Proben von Patient E	230
F.6	Entwicklung des phänotypischen Resistenzprofils der Proben von Patient F	232
F.7	Entwicklung des phänotypischen Resistenzprofils der Proben von Patient G	234
F.8	Entwicklung des phänotypischen Resistenzprofils der Proben von Patient H	236
F.9	Entwicklung des phänotypischen Resistenzprofils der Proben von Patient I	239
F.10	Entwicklung des phänotypischen Resistenzprofils der Proben von Patient J	241

F.11 Entwicklung des phänotypischen Resistenzprofils der Proben von Patient K	243
F.12 Entwicklung des phänotypischen Resistenzprofils der Proben von Patient L	245
F.13 Entwicklung des phänotypischen Resistenzprofils der Proben von Patient M	248
G.1 Ergebnisse des Mischungsdetektors für die nachsequenzierten Proben	251

Liste der Quelltexte

2.1	Drei NGS-Sequenzfragmente im FastQ-Format.	29
2.2	Minimalbeispiel zweier Objekte im JSON-Format.	34
5.1	Aufruf des DBSCAN-Algorithmus mit R Version 3.0.2.	73
5.2	Aufruf des exakten binomischen Tests mit R Version 3.0.2	99
5.3	Berechnung der Fehlerwahrscheinlichkeit anhand des Basenqualitätswerts mit R Version 3.0.2	99

Abkürzungsverzeichnis

A	Adenin
AMK	Amikacin
ASCII	Amerikanischen Standardcode zum Informationsaustausch; engl. American Standard Code for Information Interchange
BAM	Binäres Sequenz-Ausrichtungs-Format
BIGsdb	Bakterienproben-Genomsequenz Datenbank; engl. Bacterial Iso- late Genome Sequence database
BLAST	Datenbanksequenzen Abgleichprogramm; engl. Basic Local Ali- gnment Search Tool
bp	Basenpaar
BWA	Burrows-Wheeler Ausrichtung; engl. Burrows-Wheeler Alignment
C	Cytosin
CASTB	The Comprehensive Analysis Server for the <i>Mycobacterium tu- berculosis</i> complex
CFX	Ciprofloxacin
CPM	Capreomycin
CSV	Komma-getrennte Werte; engl. comma-separated values
DBSCAN	Dichte-basierter Clusteralgorithmus von Daten mit Rauschen; engl. Density Based Spatial Clustering of Applications with Noi- se
DCS	D-Cycloserin
DNS	Desoxyribonukleinsäure
DR	engl. direct repeat
DST	Resistenztestung; engl. Drug Susceptibility Testing
EAI	Ostafrikanisch/Indischer Ozean; engl. East African/Indian Ocean
EMB	Ethambutol

emPCR	Emulsions-Polymerasekettenreaktion
ENA	Europäischen Nukleotid Archiv; engl. The European Nucleotide Archive
ETH	Ethionamid
FQ	Fluorochinolone
G	Guanin
GATK	Genomanalyse-Softwarepaket; engl. Genome Analysis Toolkit
GB	Gigabyte
GPL	GNU General Public License
HDI	Index der humanen Entwicklung; engl. Human Development Index
HIV	Humanes Immundefizienz-Virus; engl. human immunodeficiency virus
Indel	Insertion oder Deletion
INH	Isoniazid
IT	Informationstechnik
JSON	JavaScript Object Notation
KAN	Kanamycin
LAM	Latein-Amerikanischer Mittelmehrraum; engl. Latin-American Mediterranean
LPA	Sondentest; engl. Line Probe Assay
LSP	Lange Sequenzpolymorphismen
LZD	Linezolid
MB	Megabyte
MDR	multiresistent; engl.:multi-drug resistant
MIRU-VNTR	Mycobacterial Interspersed Repetitive Units - Variable Number of Tandem Repeats
MOX	Moxifloxacin
MTBK	<i>Mycobacterium tuberculosis</i> -Komplex
NGS	Sequenziertechniken der nächsten Generation; engl. Next-Generation Sequencing
NTM	Nichttuberkulösen Mykobakterien

OFX	Ofloxacin
PAS	Paraaminosalicylsäure
PCR	Polymerasekettenreaktion; engl. polymerase chain reaction
PhyResSE	Phylo-Resistance Search Engine
PZA	Pyrazinamid
QRDR	Quinolon-Resistenz assoziierte Kernregion, engl. Quinolone Resistance Determining Region
RFLP	Restriktions-Fragmentlängen Polymorphismus
RMP	Rifampicin
RNS	Ribonukleinsäure
RR-TB	Rifampicin-resistente TB
RRDR	RMP-Resistenz assoziierte Kernregion, engl. Rifampicin Resistance Determining Region
rRNS	ribosomale Ribonukleinsäure
SAM	Sequenz-Ausrichtungs Format; engl. Sequence Alignment/Map-Format
SM	Streptomycin
SMRT	Einzelmolekül-Echtzeit Sequenzierung; engl.: Single-Molecule Real-Time Sequencing
SNP	Einzelnukleotidaustausch; engl. Single Nucleotide Polymorphism
T	Thymin
TB	Tuberkulose
TBDReaMDB	Tuberkulose Resistenz-Datenbank; engl. Tuberculosis Drug Resistance Mutation Database
VCF	Varianten-Format; engl. Variant Call Format
WHO	Weltgesundheitsorganisation; engl.: World Health Organization
XDR	extensiv resistent; engl.: extensively drug resistant

Teil I

Einführung

1 Tuberkulose

Erstmals seit Mitte des
20. Jahrhunderts erleben wir in
Deutschland eine Umkehr des
bislang rückläufigen
Tuberkulose Trends.

Lena Fiebig, 2016

Robert Koch Institut, Berlin

Die Tuberkulose (TB) zählt auch heute noch neben Malaria und der Infektion mit dem Humanen Immundefizienz-Virus (engl.: human immunodeficiency virus; HIV) zu den zehn häufigsten zum Tode führenden Infektionskrankheiten weltweit [213].

Diese Arbeit soll einen Beitrag zur Entwicklung einer personalisierten Therapie von Tuberkulosepatienten leisten.

1.1 Geschichte der Tuberkulose

Die TB-Erkrankung begleitet den Menschen seit Jahrtausenden. Bereits in 500 000 Jahre alten Schädelknochen des *Homo erectus* konnten TB-assoziierte Krankheitsbilder nachgewiesen werden [96]. Als ältester gesicherter Nachweis gilt der von Hershkovitz *et al.* beschriebene, circa 9 000 Jahre alte, Skelettfund in Israel [78], der eine Jahrzehnte lange Koevolution des TB-Erregers mit dem Menschen bestätigt.

Bereits Aristoteles vermutete eine Übertragbarkeit der Krankheit [71]. Doch erst Mitte des 19. Jahrhunderts konnte der als „Tuberkelbazillus“ beschriebene Erreger, *Mycobacterium tuberculosis*, von Robert Koch mikroskopisch nachgewiesen werden [106] und eine Übertragung anhand von Tierversuchen belegt werden [206]. In den darauf folgenden Jahren entwickelten Franz H. P. Ziehl und

Friedrich C. A. Neelsen, die nach ihnen benannte Färbemethode (Ziehl-Neelsen Färbung) [147], die auch heute noch zum mikroskopischen Nachweis von Mykobakterien in der Diagnostik angewendet wird [97]. Mit der Entwicklung des Tuberkulin Hauttests durch Charles Mantoux und Felix Mendel stand 1912 eine weitere diagnostische Möglichkeit zur Verfügung [146].

Schon früh wurde erkannt, dass eine Infektion nur mit Hilfe einer vorbeugenden Impfung zu verhindern ist. Doch erst 1910 gelang es Albert Calmette und Camille Guerin den nach ihnen benannten BCG-Impfstoff (Bacillus Calmette-Guerin) zu entwickeln [110]. Es handelt sich um einen attenuierten Stamm, der ursprünglich (1901) aus einer Kuh (humanpathogenen TB-Erreger *M. bovis*) mit tuberkulöser Mastitis isoliert wurde. Mittlerweile ist gezeigt worden, dass der BCG-Impfstoff bei Erwachsenen nicht zuverlässig vor einer aktiven TB schützt und die Wirkung nur auf bestimmten Formen der TB, zum Beispiel die tuberkulöse Hirnhautentzündung bei Kleinkindern, beschränkt ist [31]. Aus diesem Grund wird diese Impfung seit 1998 von der Ständigen Impfkommision (STIKO) beim Robert-Koch-Institut nicht mehr empfohlen [109].

Bis zur Mitte des 20. Jahrhunderts standen gezielt auf den TB-Erreger wirkende Medikamenten nicht zur Verfügung. Zur Behandlung wurden daher natürliche Methoden eingeführt, wie zum Beispiel die von Hermann Brehmer 1856 begründete Freiluft-Liegekur in speziellen TB-Sanatorien [19], die neben der Pneumothorax das Haupttherapeutikum im letzten Viertel des 19. und Anfang des 20. Jahrhunderts darstellte [152]. Brehmer forderte bei dieser Liegekur eine lang-dauernde Behandlung an „immunen Orten“, die unterstützt wird durch Bewegung an frischer Luft und einer reichhaltigen Ernährung unter anderem mit Milch und Weinbrand [19, 152]. Mit Streptomycin (SM) stand 1944 erstmalig ein wirksames Antibiotikum gegen die TB zur Verfügung [177]. In den Folgejahren wurden weitere Antibiotika entdeckt, die in Tabelle 1.1 zusammengefasst sind. Insbesondere die Markteinführung von Isoniazid (INH) und Rifampicin (RMP) verbesserten die Therapie erheblich [35, 107].

Neben den nun zur Verfügung stehenden wirksamen Medikamenten führten auch verbesserte Lebensumstände zu einem stetigen Rückgang der Fallzahlen bis etwa Mitte der 1980er Jahre [136]. Dann war ein erheblicher Anstieg von TB-Fällen zu verzeichnen [70]. Die Entwicklung war so dramatisch, dass die Weltgesundheitsorganisation (engl.: World Health Organization, WHO) 1993 den globalen TB-Notstand ausgerufen hat [70]. Experten machten den Anstieg der Weltbevöl-

kerung, Armut, Migration, Reduktion von TB-Kontrollprogrammen, aber auch vor allem die Koinfektion mit HIV und die Ausbreitung von resistenten TB-Stämmen dafür verantwortlich [39].

Im Jahr 1998 wurde von der WHO, zusammen mit einer Vielzahl von internationalen, öffentlichen und privaten Organisationen die Partnerschaftsgemeinschaft „Stop TB Partnership“ gegründet. Die Hauptziele des von der Partnerschaftsgemeinschaft entwickelten globalen Plans zur Eindämmung der TB (engl.: Global Plan to Stop TB, [159]) waren die Halbierung der TB-Prävalenz und TB-Sterblichkeit bis 2015 im Vergleich zum Jahr 1990 [159]. Basierend auf den Ergebnissen des globalen Plans wurde im Mai 2014 ein weiterführender Plan (engl.: Global Plan to End TB) für die Jahre 2015-2035 von der WHO verabschiedet [155]. Zu den Meilensteinen gehören die Reduzierung der TB-Todesfälle um 95 % sowie die Reduzierung der Anzahl neuer TB-Fälle um 90 % zwischen den Jahren 2015 und 2035 [155]. Das Herzstück der globalen Pläne ist die Einhaltung und Verfügbarkeit der so genannte „DOTS-Strategie“ (engl.: Directly Observed Treatment Short Course), eine von der WHO in den 1990er Jahren eingeführte Kontrollstrategie [154]. Diese soll die Heilungschancen verbessern und der Resistenzentwicklung entgegenwirken. Die kontrollierte Medikamenteneinnahme ist ein Bestandteil dieser Strategie. Darüber hinaus sollen Regierungen und Gesundheitsministerien die TB-Bekämpfung mit oberster Priorität behandeln und entsprechende Ressourcen zur Verfügung stellen. Dazu gehört eine qualitätskontrollierte Fallfindung und Diagnose für ein standardisiertes Berichtswesen durchzuführen und die Medikamentenversorgung für jeden Patienten zu jeder Zeit sicher zu stellen [154].

1.2 Epidemiologie der Tuberkulose

Trotz intensiver Bemühungen konnten die Meilensteine der Globalen Pläne zur Eindämmung der TB der „Stop TB Partnership“ (Kapitel 1.1) bis jetzt nicht erreicht werden [212]. Nach Schätzungen der WHO ist etwa ein Drittel der Weltbevölkerung mit dem TB-Erreger latent infiziert. Schätzungsweise entwickeln 5-15 % der Infizierten im Laufe ihres Lebens eine aktive, beziehungsweise offene TB [152, 213].

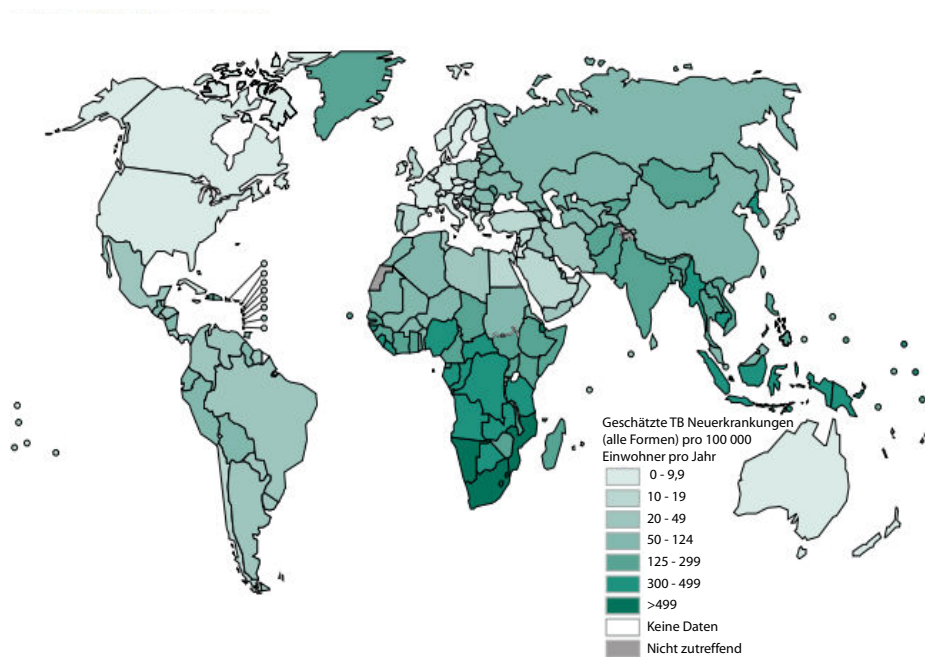


Abbildung 1.1: Geschätzte TB Neuerkrankungen (alle Formen) pro 100 000 Einwohner für die Regionen der Welt im Jahr 2015. Besonders hohe Raten können in Indien, Indonesien, China, Nigeria, Pakistan und Südafrika beobachtet werden. Quelle: [213]

Insgesamt gab es im Jahr 2015 laut WHO 10,4 Millionen TB-Neuerkrankungen und 1,4 Millionen Todesfälle [213]. Damit ist die TB auch heute noch unter den behandelbaren Infektionskrankheiten, diejenige, die weltweit die meisten Todesopfer fordert [171].

Abbildung 1.1 zeigt die TB-Inzidenz (Neuerkrankungen/100 000 Einwohner in einem bestimmten Zeitraum) für die verschiedenen Regionen der Welt im Jahr 2015. Sechzig Prozent der weltweiten Neuerkrankungen fallen dabei allein auf die Länder Indien, Indonesien, China, Nigeria, Pakistan und Südafrika [213].

Die TB-Kontrolle wird durch die Koinfektion mit HIV erschwert. Patienten mit HIV haben durch die Schwächung des Immunsystems ein 20-30-fach erhöhtes Risiko nach einer TB-Infektion eine aktive TB zu entwickeln, als TB-Infizierte ohne eine Koinfektion mit HIV, wodurch HIV-Patienten schneller infektiös sind und die TB-Erreger übertragen können [150]. Im Jahr 2015 entwickelten 1,2 Millionen HIV-Infizierte zusätzlich eine TB, von denen 400 000 an dieser verstarben [213].

Ein zunehmendes Problem bei der Behandlung von TB stellt das vermehrte

Auftreten von resistenten Stämmen dar [152]. Insgesamt erkrankten im Jahr 2015 480 000 Patienten an einer so genannten multiresistenten Form der TB (engl.: multi-drug resistant, MDR) [213]. Dabei handelt es sich um eine Erkrankung mit einem Erreger-Stamm, der definitionsgemäß Resistenzen gegenüber den zwei effektivsten Erstrangmedikamenten INH und RMP aufweist. Weitere 100 000 Patienten entwickelten im Jahr 2015 eine RMP-resistente TB (RR-TB), die eine zur MDR-TB vergleichbare Behandlung indiziert. In Abbildung 1.2 ist der Anteil von MDR/RR-TB-Fällen an allen detektierten TB-Fällen im Jahr 2015 für die Regionen der Welt dargestellt. Die Länder mit der größten MDR/RR-TB Last, basierend auf den absoluten Zahlen aus dem Jahr 2015, sind Indien mit 130 000 Fällen, China mit 70 000 Fällen und die Russische Föderation mit 60 000 Fällen (gemeinsam insgesamt 45 % der globalen MDR/RR-TB-Fälle) [213]. Für Russland bedeutet das zum Beispiel, dass mehr als 18 % der TB-Neuerkrankungen MDR/RR-TB-Fälle sind (Abbildung 1.2).

Eine noch größere Herausforderung für die TB-Kontrolle stellt das Vorkommen der, erstmals 2006 beschriebenen, extrem resistenten TB (engl.: extensively drug resistant, XDR) dar. Bei dieser Form handelt es sich um die Infektion mit einem Erreger, der neben INH und RMP zusätzlich gegen mindestens einem der Fluorchinolone (FQs) und einem der drei injizierbaren Zweitrangantibiotika Amikacin (AMK), Kanamycin (KAN) oder Capreomycin (CPM) resistent ist [140, 186]. Mittlerweile wurden XDR-TB-Fälle in 117 der insgesamt 194 Mitgliedstaaten der WHO beschrieben. In 2015 wurden 7 579 Patienten aus 74 verschiedenen Ländern mit einer XDR-TB registriert. Der Großteil der Fälle wurde in Indien (2 130), der Ukraine (1 206), der Russischen Föderation (1 205) und Südafrika (719) detektiert [213].

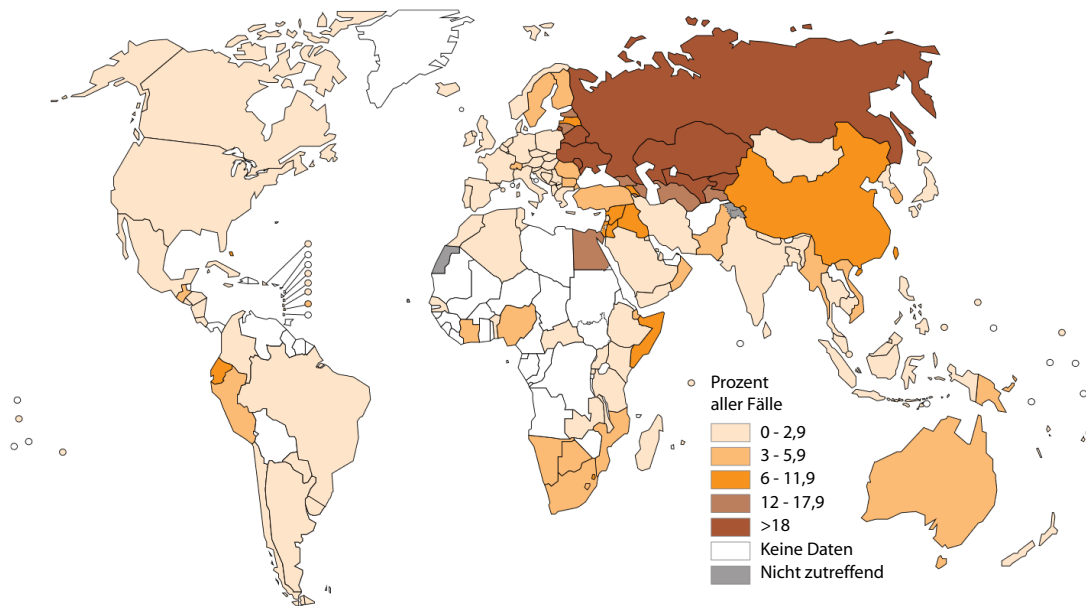


Abbildung 1.2: Anteil (in %) neuer MDR/RR-TB-Fälle an allen TB-Fällen pro Region im Jahr 2015. Quelle: [213]

1.3 Die Tuberkuloseinfektion

Die Erreger der TB werden hauptsächlich von Erkrankten mit einer so genannten offenen TB durch Tröpfcheninfektion übertragen [98]. Die Erreger gelangen in die Lunge und es kommt zu einer unspezifischen Entzündungsreaktion. Die Erreger werden von den Alveolarmakrophagen aufgenommen und teilweise verdaut [176]. Allerdings besitzen die Erreger der TB spezielle Abwehrmechanismen, welche ihnen ein Überleben und sogar die Vermehrung in den Makrophagen ermöglichen [176]. Die infizierten Makrophagen rekrutieren weitere Immunabwehrzellen und mit Hilfe der T-Zellen bildet sich eine für die Infektion mit dem Erreger typische granulomatöse Entzündungsreaktion [182]. Dabei produziert der Körper einen Zellwall um die infizierten Makrophagen, den so genannten Lungenherd oder auch Granulom [98]. Auf einem Thorax-Röntgenbild lassen sich diese knötchenförmigen Gewebeneubildungen nachweisen [202]. Auf diese Weise kontrolliert das menschliche Immunsystem die Bakterien und verhindert eine Ausbreitung (latente TB). Solange das Immunsystem intakt ist, kann ein Ausbruch verhindert werden [176].

Auslöser für den Ausbruch der Krankheit (aktive, offene TB) ist unter anderem die Schwächung der zellulären Antwort des Immunsystems durch beispielsweise Unterernährung, mangelnde Hygiene, Einnahme von Immunsuppressiva oder andere Infektionen, wie zum Beispiel eine Infektion mit HIV [44, 123]. Zu den Symptomen der aktiven TB gehören zum Beispiel anhaltender Husten, Müdigkeit, Gewichtsverlust, Nachtschweiß und leichtes Fieber [44].

Die Entzündungsreaktion führt im Endstadium durch nekrotisierende und fibrosierende Prozesse zur Gewebszerstörung [94]. Unbehandelt nimmt die TB bei etwa 50 % der Infizierten innerhalb von zwei Jahren einen tödlichen Verlauf [44].

Ähnlich schlechte Heilungschancen bestehen trotz adäquater Antibiotika-Therapie für Patienten mit einer MDR-TB. Nur etwa 50-70 % der MDR-TB-Patienten werden in Abhängigkeit von der Region erfolgreich therapiert [114, 157].

1.4 *M. tuberculosis*-Komplex

Die Erreger der TB gehören zur Familie der *Mycobacteriaceae*. Bei den Erregern aus dem *M. tuberculosis*-Komplex (MTBK) handelt es sich um ein 1-4 μm langes

unbewegliches, säurefestes Stäbchenbakterium mit einem Umfang von 0,3 - 0,6 μm [187]. Der Begriff säurefest geht auf den Nachweis von Mykobakterien über die Ziehl-Neelsen Färbung zurück. Dabei werden Zellen mit Karbolfuchsin gefärbt, anschließend durch eine Säurebehandlung entfärbt und mit Methylenblau behandelt [99]. Mykobakterien erscheinen im Gegensatz zu allen anderen Methylenblau-gefärbten Bakterien rot, da sie aufgrund des besonderen Aufbaus ihrer Zellwand durch die Säurebehandlung nicht entfärbt werden [99]. Die mykobakterielle Zellwand ist unter anderem gekennzeichnet durch die namengebenden Mykolsäuren, dem so genannten Cord-Faktor (Glykolipid: 6,6'-Dimykolytrehalose) und Wachsen [89]. Der komplexe Aufbau ermöglichen den Bakterien nicht nur das Überleben und die Vermehrung in Makrophagen durch die Hemmung der Phagosom-Lysosom Fusion [11, 191], sondern verringert zusätzlich die Permeabilität für viele hydrophile Antibiotika, wie die Beta-Lactam-Antibiotika (z. B. Penicillin), wodurch die TB-Erreger ein hohes Maß an intrinsischer Resistenz aufweisen [20, 89]. *M. tuberculosis* ist ursächlich für 95% der TB-Erkrankungen beim Menschen [153]. Weitere Erreger sind zum Beispiel *M. bovis* und *M. pinnipedii*, die zusammen mit *M. tuberculosis* und weiteren Spezies im MTBK zusammengefasst werden (Kapitel 1.5).

1.5 Populationsstruktur des MTBK

Die Erreger der verschiedenen Spezies des MTBK unterscheiden sich in ihrem Wirtsspektrum und der regionalen Häufigkeit ihres Auftretens [82]. So sind die Erreger der Spezies *M. tuberculosis*, *M. africanum* und *M. canettii* hauptsächlich humane Pathogene, während zum Beispiel die Erreger der Spezies *M. bovis* (Rinder), *M. caprae* (Ziegen), *M. microti* (Nagetiere) und *M. pinnipedii* (Seehunde) vorwiegend Tiere als Wirt haben, die namengebend für die jeweilige Spezies sind [82].

Basierend auf molekularbiologischen Typisierungsmethoden lassen sich die MTBK-Stämme in verschiedene phylogenetische Linien unterteilen, die in Abbildung 1.3 dargestellt sind. Die phylogenetischen Linien sind vielfach nach der Region benannt worden, in der sie zuerst beschrieben wurden [211]. Gagneux *et al.* führte erstmalig 2006 ein numerisches Benennungssystem ein, in dem sieben Hauptlinien, zum Beispiel Linie 2 (Linie Beijing) oder Linie 4 (Euro-Amerikanische Superlinie),

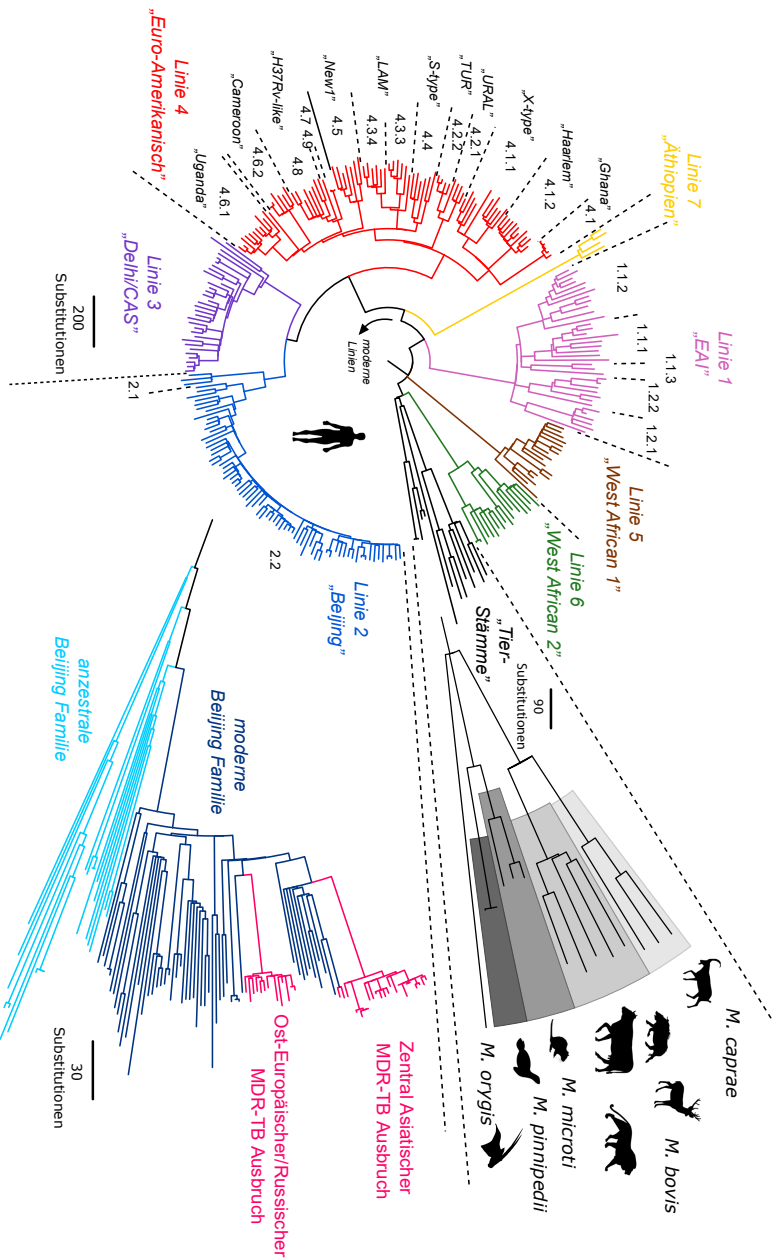


Abbildung 1.3: Neighbour-Joining Baum zur Darstellung der globalen phylogenetischen Struktur des *M. tuberculosis*-Komplexes (erzeugt aus 1 000 Bootstrap-Replikaten basierend auf 36 963 SNPs von 286 MTBK Stämmen). MTBK-Isolate können in sieben Hauptlinien (farblich abgegrenzt) und mehrere Untergruppen klassifiziert werden. Die so genannten „modernen“ Linien des MTB (Linien 2, 3, 4) kommen weltweit vor, während Infektionen mit Stämmen der „anzestralen“ Linien des MTBK hauptsächlich in West und Ost Afrika vorkommen. Quelle: [150]

unterschieden werden (Abbildung 1.3) [17, 65]. Wie in Abbildung 1.3 gezeigt ist, korrelieren beide Klassifizierungssysteme sehr gut und können analog verwendet werden.

Zu den klassischen Typisierungsmethoden zählen neben der *IS6110* (Insertionssequenz *6110*) Desoxyribonukleinsäure (DNS) Fingerprint Typisierung, die Spacer Oligonukleotid (Spoligotyping) sowie die MIRU-VNTR („Mycobacterial Interspersed Repetitive Units - Variable Number of Tandem Repeats“) Typisierung.

- *IS6110* DNS Fingerprint Typisierung [50]:
Diese Methode basiert auf der Detektion der Insertionssequenz *IS6110*. *IS6110* ist eine repetitive, mobile Sequenz, die in variabler Kopienzahl im Genom von MTBK-Stämmen vorkommt [161]. Für die Analyse wird die genomische DNS mittels des Restriktionsenzym *PvuII* fragmentiert [131]. Anschließend werden die entstandenen Fragmente gelelektrophoretisch der Größe nach aufgetrennt, auf eine Nylonmembran transferiert und mit Hilfe einer markierten Sonde, die komplementär zum 3'-Ende der *IS6110* ist, nachgewiesen [131]. Die Kopienzahl und Genompositionen der *IS6110* Elemente variieren, sodass sich auf diese Weise für jeden Stamm ein spezifisches Bandenmuster ergibt [50].
- Spacer Oligonukleotid Typisierung (Spoligotypisierung) [95]:
Dieses Verfahren untersucht das Vorhandensein beziehungsweise Fehlen einzigartiger Spacer DNS Sequenzen in repetitiven Genomregionen (engl. direct repeat (DR) region). Insgesamt werden 43 dieser Oligonukleotid-Spacer verwendet [95]. Die einzelnen phylogenetischen Linien unterscheiden sich in dem Vorkommen der 43 Sequenzen und dem daraus resultierenden Typisierungsmuster.
- MIRU-VNTR Typisierung [63, 132, 199]:
Bei dieser Methode wird die Anzahl der Wiederholungen (Kopienzahl) bestimmter kurzer Sequenzen in repetitiven Genomregionen, so genannten Mikro- beziehungsweise Minisatelliten, bestimmt (engl.: Mycobacterial Interspersed Repetitive Units). Die entsprechenden Regionen werden mit Hilfe der Polymerase-Kettenreaktion (engl.: polymerase chain reaction, PCR) amplifiziert und im Anschluss der Größe nach aufgetrennt. Auf diese Weise kann die jeweilige Kopienzahl der MIRU-VNTR-Marker bestimmt werden. Es werden standardisiert 24 dieser Regionen in die Analyse einbezogen (24

Loci-MIRU-VNTR), um eine möglichst hohe Auflösung der einzelnen Linien zu erreichen. Jeder Stamm wird demnach durch einen spezifischen 24-stelligen Zahlencode charakterisiert, der sich aus der Kopienzahl der MIRU-VNTR-Marker ergibt.

Neben den klassischen Typisierungsmethoden gewinnen genombasierte Verfahren, wie zum Beispiel die Klassifizierung anhand spezifischer Einzelnukleotidaustausche (engl.: Single Nucleotide Polymorphisms, SNPs), immer mehr an Bedeutung [32, 82]. Die Studie von Coll *et al.* [32] untersuchte 1601 MTBK-Genome aus elf unabhängigen Sequenzierstudien aus verschiedenen Regionen der Welt. In diesem Datensatz waren Stämme aller sieben Hauptlinien (Linie 1: n = 121, Linie 2: n = 390, Linie 3: n = 189, Linie 4: n = 856, Linie 5: n = 17, Linie 6: n = 11, Linie 7: n = 6) sowie MTBK-Stämme der Spezies *M. bovis* (n = 11) vertreten. Basierend auf der vergleichenden Analyse der Genomdaten, konnten 62 SNPs identifiziert werden, mit denen die sieben Hauptlinien und 55 Untergruppen, zum Beispiel Linie 4.3.3 (Linie LAM) oder Linie 1.1.1 (Linie EAI), eindeutig unterschieden werden können (Abbildung 1.3). Insgesamt ist mit genombasierten Methoden eine höhere Präzision als mit den klassischen Typisierungsmethoden zu erreichen [108, 158]. So ist eine Abgrenzung von Stämmen, die beispielsweise mit dem gleichen MIRU-VNTR Typisierungsmuster klassifiziert wurden, mit Hilfe der Gesamtgenom-Analyse bei Ausbruchsuntersuchungen anhand weniger SNP-Unterschiede möglich [108].

Genetische Unterschiede zwischen verschiedenen phylogenetischen Linien sind auch von großer Bedeutung hinsichtlich der Virulenz, der Übertragbarkeit und der Resistenzentwicklung [77, 164]. So spielen Veränderungen im Erbgut (SNPs, Insertionen/Deletionen [Indels]), die in Zielstrukturen der Antibiotika lokalisiert sind (Tabelle 1.1) oder die zu Veränderungen in Enzymen führen, die zur Aktivierung von inaktiven Wirkstoffen, so genannten Prodrugs (z. B.: Pyrazinamid, PZA), benötigt werden, eine bedeutende Rolle im Infektionsverlauf und in der Therapie [72]. Beispielhaft untersuchte eine Studie aus dem Jahr 2013 die Bedeutung des genetischen Hintergrunds bei der Ausbildung von Antibiotikaresistenzen [61]. Ford *et al.* beobachteten bei Stämmen der Linie Beijing (Linie 2) *in vitro* einen schnelleren Resistenzwerb als bei Stämmen der Euro-Amerikanischen Superlinie (Linie 4). Insbesondere zeigte sich für Stämme der Linie 2 im Vergleich zu Stämmen der Linie 4 eine höhere Mutationsrate für das Gen *rpoB*, das Resistenz gegenüber RMP vermittelt (Tabelle 1.1). Aber nicht nur für RMP, sondern auch

für INH und Ethambutol (EMB) konnte bei Stämmen der Linie 2 im Vergleich zu Stämmen der Linie 4 ein schnellerer Resistenzerwerb beobachtet werden. Das Risiko der Ausprägung einer MDR-TB (INH- und RMP-Resistenz) wäre für Träger der Linie Beijing (Linie 2) nach dieser Studie höher als bei der Infektion mit Stämmen der Euro-Amerikanischen Superlinie (Linie 4) [61].

1.6 Diagnose und Therapie der Tuberkulose

Eine effektive Therapie, die auf einer frühen und validen Diagnostik aufbaut, ist von größter Bedeutung für die TB-Kontrolle und Behandlung [212]. Neben dem Nachweis einer aktiven TB-Infektion, zum Beispiel durch mikroskopische Untersuchung des Sputums mit Hilfe zum Beispiel der Ziehl-Neelsen Färbung oder der Anzucht auf speziellen Nährmedien [190], ist die Untersuchung auf vorliegende Resistenzen unverzichtbar für die Einleitung einer effektiven Therapie [159]. Die Resistenztestung (engl.: Drug Susceptibility Testing, DST) kann mit phänotypischen, wachstumsbasierten oder molekularbiologischen Verfahren durchgeführt werden [190].

Grundsätzlicher Nachteil der Kulturverfahren ist das langsame Wachstum (Generationszeit 24-48 Stunden) der TB-Erreger. So dauert der Nachweis von Resistenzen auf Festmedium acht bis zwölf Wochen und in Flüssigmedium 7-13 Tage [3, 187].

Darüber hinaus sind phänotypische DST-Ergebnisse zum Teil schwer reproduzierbar [83] und mehrere Studien berichten für einige Erstrangantibiotika, wie zum Beispiel EMB, nur geringe Sensitivität (83,3 % [95 % Konfidenzintervall, 42,0-97,2]) und Spezifität (83,9 % [95 % Konfidenzintervall, 72,7-91,1]) [83, 102].

Zu den molekulargenetischen Verfahren der Resistenztestung gehören neben den DNS-Sondentests, bei denen eine Probe zum Beispiel auf spezifische Sonden hybridisiert wird, die Detektion von Mutationen mittels Sequenzierung (Kapitel 1.8). Ein Vorteil gegenüber phänotypischen Verfahren ist die Nachweiszeit, so können zum Beispiel mit Hilfe des Xpert[®] MTB/RIF-Verfahrens der Firma Cepheid innerhalb von zwei Stunden Erreger des MTBKs nicht nur nachgewiesen werden, sondern auf eine RMP-Resistenz getestet werden [129, 193]. Eine RMP-Resistenz wird als Indiz für eine vorliegende MDR-TB verwendet und leitet eine zur MDR-

Therapie ähnliche Behandlung ein [193]. Allerdings fehlen für alle weiteren Erst- und Zweitrangmedikamente Informationen über vorliegenden Resistenzen; eine individualisierte Therapie ist daher erst nach Durchführung von Sondentests (engl.: Line Probe Assays, LPA) oder phänotypischer DST möglich [5].

Die für die TB-Behandlung verwendeten Antibiotika wirken auf unterschiedliche Weise. Zu den bakterizid, also abtötend wirkenden Antibiotika zählen die Erstrangmedikamente INH, RMP, PZA sowie einige Zweitrangmedikamente (Tabelle 1.1). Als bakteriostatisch, also wachstumshemmend, gilt EMB (Tabelle 1.1). Darüber hinaus gelten PZA und RMP als sterilisierende Antibiotika, die zusätzlich auf persistierende Bakterien wirken und die Behandlungsdauer dadurch signifikant verringern [110]. Die derzeit von der WHO empfohlene Behandlung besteht aus einer Kombination verschiedener Antibiotika mit unterschiedlichen Wirkmechanismen (Tabelle 1.1), sodass Bakterien unterschiedlicher physiologischer Stadien (latent/aktiv) abgetötet werden und die Wahrscheinlichkeit einer Selektion resistenter Zellen minimiert wird [171, 210].

Nach Vorgaben der WHO werden Patienten mit einer sensiblen TB über einen Zeitraum von insgesamt sechs Monaten mit vier beziehungsweise zwei Antituberkulotika behandelt [210]. Die Standardmedikamente (Erstrangantibiotika) sind INH, RMP, EMB und PZA. In der initialen Phase (zwei Monate) erfolgt eine tägliche Behandlung mit den vier Erstrangmedikamenten INH, RMP, PZA und EMB. In der anschließenden Stabilisierungsphase (vier Monate) werden nur noch die zwei effektivsten Antibiotika INH und RMP verabreicht [210].

Durch eine fehlende Kombinationstherapie oder durch geringe Therapietreue des Patienten während der Behandlung können resistente Pathogene selektioniert werden [100, 205]. Resistenzvermittelnde Variationen entstehen spontan und zufällig in einer Bakterienzelle, wobei die Mutationsrate in Abhängigkeit des betrachteten Lokus zwischen 10^{-6} und 10^{-8} Mutationen pro Bakterium und Generation liegt (Vergleich *E. coli*: $10^{-3} - 10^{-4}$) [12, 133, 144]. Das bedeutet, in einer Population mit $1 * 10^8$ Bakterien und einer Mutationsrate von $1 * 10^{-8}$ trägt im Durchschnitt eines eine Mutation in einem resistenzvermittelnden Gen. Unter antibiotischen Selektionsdruck, werden diese selektiert und es kommt zur Bildung einer resistenten Subpopulation, die sich während der Behandlung in der Gesamtpopulation durchsetzt [100].

Bei Vorliegen von Resistenzen, aber auch bei Unverträglichkeit von Erstrangmedikamenten, wird auf die Zweitrangmedikamente zurückgegriffen, zu denen

unter anderem die Fluorochinolone Ciprofloxacin (CFX), Moxifloxacin (MOX) und Ofloxacin (OFX), die injizierbaren Antibiotika AMK, CPM und KAN sowie Ethionamid (ETH), D-Cycloserin (DCS), Linezolid (LZD) zählen (Auswahl der wichtigsten Antituberkulotika in Tabelle 1.1). Insbesondere bei Patienten mit einer MDR-TB verlängert sich dabei der Behandlungszeitraum deutlich auf mindestens 20 Monate [156]. Darüber hinaus geht die Therapie unter Einsatz von Zweitrangmedikamenten mit dem Auftreten von starken Nebenwirkungen einher [45]. Dazu gehören unter anderem Nervenentzündungen, die Einschränkung des Farbsehens, der Hörverlust und die Beeinträchtigung der Leberfunktion. Die enormen Kosten (58 € pro Tag bei 600 Tagen Behandlungszeitraum verglichen mit 4 € pro Tag bei 180 Tagen Behandlungszeitraum einer sensiblen TB), die durch die Behandlung einer resistenten TB entstehen, stellen eine große Herausforderung für die Gesundheitssysteme dar [45].

Tabelle 1.1: Die wichtigsten Antituberkulotika und ihre Wirkmechanismen.

	Antibiotikum	Abk.	Markteinführung	Wirkungsweise	Wirkungsort	Resistenzgene	Beteiligte Proteine
Erstrang	Isoniazid	INH	1952	bakterizid	Fettsäurebiosynthese (Zellwand)	<i>katG</i> <i>inhA</i> <i>ahpC</i>	Katalase-Peroxidase Enoyl-ACP-Reduktase Alkyl Hydroperoxidase Reduktase
	Rifampicin	RMP	1966	bakterizid, sterilisierend	RNS-Polymerase (Transkription)	<i>rpoB</i>	β -Untereinheit der RNS-Polymerase
	Pyrazinamid	PZA	1952	bakterizid, sterilisierend	trans-Translation	<i>pncA</i>	Pyrazinamidase
	Ethambutol	EMB	1961	bakteriostatisch	Stoffwechsel	<i>embB</i>	Arabinosyltransferasen
Zweitrang	Streptomycin	SM	1944	bakterizid	Proteinbiosynthese	<i>rrs</i> <i>rpsL</i>	16S rRNS Ribosomales Protein S12
	Fluoroquinolone (CFX, MOX, OFX, u.a.)	FQ	1963	bakterizid	Topologie der DNS-Moleküle	<i>gyrA</i> <i>gyrB</i>	Gyrase Untereinheit A Gyrase Untereinheit B
	Ethionamid	ETH	1956	bakterizid	Fettsäurebiosynthese (Zellwand)	<i>inhA</i>	Enoyl-ACP-Reduktase
	Amikacin	AMK	1957	bakterizid	Proteinbiosynthese	<i>rrs</i>	rRNS Synthese
	Kanamycin	KAN	1957	bakterizid	Proteinbiosynthese	<i>rrs</i>	rRNS Synthese
	Capreomycin	CPM	1960	bakterizid	Proteinbiosynthese	<i>rrs</i>	rRNS Synthese
	D-Cycloserin	DCS	1975	bakteriostatisch	Zellwandsynthese	<i>cycA</i>	D-Serin/ Alanin/Glycin Transporter
Linezolid	LZD	2001	bakteriostatisch	Proteinsynthese	<i>rrl</i> <i>rplC</i>	ribosomale 50S Untereinheit Ribosomales Protein S50	

Abkürzungen: RNS-Ribonukleinsäure; rRNS-ribosomale Ribonukleinsäure

1.7 Heteroresistente und Mischpopulationen

TB-Patienten können simultan mit antibiotika-sensiblen und antibiotika-resistenten TB-Stämmen infiziert sein. Die Detektion dieser heterogene Populationen stellt ein Problem für eine valide Diagnose dar [60, 169]. Diese Situation ist für den Patienten besonders kritisch, wenn ausschließlich der sensible Stamm identifiziert und eine entsprechende Therapie initiiert wird. Da die resistenten Zellen nicht effektiv behandelt, beziehungsweise weiter selektioniert werden, tritt keine Verbesserung für den Patienten unter Therapieverlauf ein [60].

Wie gut heteroresistente Populationen mittels molekularbiologischer Verfahren diagnostiziert werden können, ist kaum untersucht worden. Folkvarsdn *et al.* verglichen 2013 die phänotypische Resistenztestung mit molekularbiologischen Verfahren (GenoType MTBDR*plus*, INNO-LIPA Rif.TB und *rpoB*- bzw. *katG*-Sequenzierung) für Suspensionen mit 1 %, 5 %, 10 %, 50 % und 100 % RMP- und INH-resistenter Bakterienzellen. In diesen Studien konnten die Proben mit der kritische Konzentrationen von 1 % resistenten Keimen ausschließlich mit der phänotypischen Resistenztestung zuverlässig detektiert werden [59, 60]. Mit dem GenoType MTBDR*plus* war es in dieser Studie möglich, Konzentrationen von 5 % resistenten Keimen zu detektieren und mit der Sequenzierung wurden die Proben mit mindestens 10 % (*katG*) beziehungsweise 50 % (*rpoB*) resistenten Keimen korrekt detektiert [59, 60]. Demnach konnten für geringe Konzentrationen an resistenten Keimen bessere Detektionswerte mit der phänotypischen Resistenztestung im Vergleich zu den molekularbiologischen Verfahren beobachtet werden [59, 60]. Generell können Heteroresistenzen durch verschiedene Situationen entstehen [30]:

1. Wirtsinterne Diversifikation unter Selektionsdruck
2. Mischinfektion verschiedener Stämme zu Krankheitsbeginn
3. Infektion mit einer zweiten Population während der Behandlung

Die wirtsinterne Diversifikation spiegelt den normalen Vorgang beim Resistenzerwerb, zum Beispiel unter nicht effektiver Therapie, wieder. Ausgehend von wenigen resistenten Zellen in einer Population, werden unter nicht effektiver Therapie die resistenten Zellen selektioniert, während die sensiblen Bakterien abgetötet werden [205]. Dadurch können an verschiedenen molekularbiologischen Resistenzmarkern Wildtyp und mutierte, resistenzvermittelnde Allele vorhanden sein [169].

Neben der wirtsinternen Diversifikation kann es ebenfalls durch simultane Infektion verschiedener Stämme (Mischinfektion, Reinfektion) zur Heteroresistenz kommen [221]. Es wurde lange Zeit davon ausgegangen, dass der menschliche Körper durch eine Infektion mit dem TB-Erreger immun gegenüber einer Neuinfektion mit einem weiteren MTBK-Stamm ist [29]. Obwohl bereits 1975 von Mankiewicz *et al.* Mischungen bei der Typisierung von MTBK-Stämmen einer Inuit-Population festgestellt wurden [126], hielt sich diese Hypothese Jahrzehnte lang. In den letzten Jahren wurden jedoch mehrere Fälle beschrieben, bei denen Patienten simultan mit zwei unterschiedlichen Stämmen infiziert waren [29, 30, 84, 125, 183, 221].

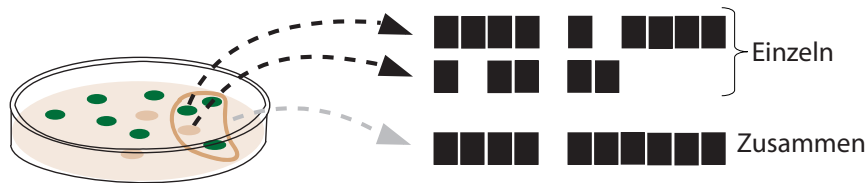
Eine Mischinfektion kann zum einen entstehen, wenn bereits bei der Aktivierung persistierender Krankheitsherde unterschiedliche Populationen durch unabhängige Mehrfachinfektion des Patienten koexistieren. Zum anderen kann eine Mischinfektion durch Reinfektion mit einem weiteren Stamm entstehen, was vor allem in Hochinzidenz-Regionen beobachtet wird [29].

Zur Heteroresistenz kann es bei Mischinfektionen dann kommen, wenn nur einer der beiden Stämme eine oder mehrere resistenzvermittelnde Mutationen trägt oder die unterschiedlichen Stämme verschiedene Resistenzprofile haben [221].

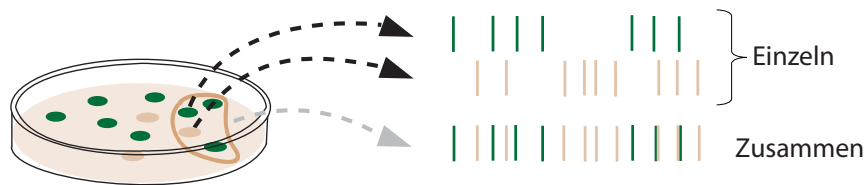
Von großer Bedeutung bei Mischinfektionen ist neben der Detektion eventuell vorhandener Heteroresistenzen zusätzlich die Identifikation der beteiligten phylogenetischen Linien, da die genetischen Unterschiede funktionelle Konsequenzen haben können (Kapitel 1.5) [77, 164]. Mit den derzeit verwendeten Typisierungsmethoden (Kapitel 1.5) ist es jedoch nur bedingt möglich, eine Mischung zweier unterschiedlicher MTBK-Stämme zu identifizieren (Abbildung 1.4) [29]. In Abbildung 1.4 sind die molekularbiologischen Typisierungsmethoden zur Detektion von Mischinfektionen zusammengefasst.

Bei der Spoligotypisierung (Abbildung 1.4a) können entweder durch Subkulturen unterschiedliche Typisierungsmuster beobachtet werden (Abbildung 1.4a) oder im Spezialfall können schwächere Intensitäten an bestimmten Spacern beobachtet werden, die auf eine Mischinfektion hindeuten [29]. Wird hingegen nur das gemeinsame Typisierungsmuster beobachtet, kann dies zur Fehlinterpretation führen. Insgesamt wird die Detektion von Mischinfektionen basierend ausschließlich auf dem Ergebnis der Spoligotypisierung nicht empfohlen [134].

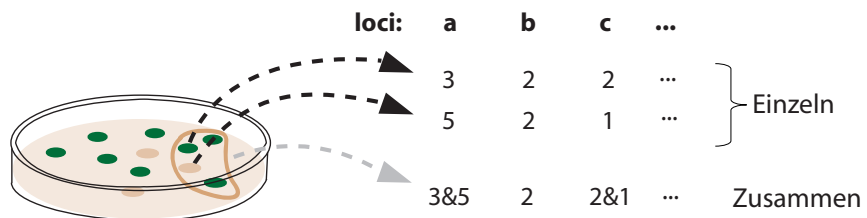
Bei der IS6110 DNS Fingerprint Typisierung können entweder durch Subkulturen verschiedene Bandenmustern nachgewiesen oder im Falle einer fehlenden



(a) Mischdetektion mit Hilfe der Spoligotypisierung



(b) Mischdetektion mit Hilfe der IS6110 DNS Fingerprint Typisierung



(c) Mischdetektion mit Hilfe der MIRU-VNTR Typisierung

Abbildung 1.4: Detektion von Mischinfektionen mit den standardisierten Typisierungsmethoden. **a** Detektion mit der Spoligotypisierung. Subkulturen zeigen unterschiedliche Bandenmuster. Eine Mischkultur beider Stämme zeigt ein gemeinsames Bandenmuster, das zu keinem der einzelnen Stämme gehört. **b** Detektion mit der IS6110 DNS Fingerprint Typisierung. Subkulturen zeigen unterschiedliche Bandenmuster. Eine Mischkultur zeigt eventuell schwächere Banden des unterrepräsentierten Stammes. **c** Detektion mit der MIRU-VNTR Typisierung. Subkulturen zeigen unterschiedliche Zahlencodes. Eine Mischkultur zeigt unterschiedliche Kopienzahlen am selben Locus. Quelle: [29]

Subkultivierung zusätzliche, schwächere Banden im selben Gel detektiert werden (Abbildung 1.4b). Das Auftreten schwächerer Banden kommt allerdings nur vor, wenn einer der beiden Stämme in der Probe unterrepräsentiert ist [29]. Darüber hinaus kann es oftmals in der Diagnostik als Artefakt fehlinterpretiert werden.

Bei der MIRU-VNTR Typisierung kann eine Mischinfektion ebenfalls an verschiedenen Zahlencodes unterschiedlicher Subkulturen derselben Probe oder an zwei verschiedenen Kopienzahlen am selben Locus identifiziert werden (Abbildung 1.4c). Die Schwierigkeit besteht darin, eine Mischinfektion von einer Allel-Evolution zu unterscheiden (Abbildung 1.4c) [29]. Mischungen nah verwandter Stämme, die zum Beispiel nur einen Locus Unterschied haben, können fehlinterpretiert werden.

Das Auftreten von Mischinfektionen ist hinsichtlich ihres Vorkommens, aber auch ihrer Häufigkeit kontrovers diskutiert. Van Rie *et al.* vermuten Mischinfektion hauptsächlich in Regionen mit hohen Inzidenzraten (Abbildung 1.1) [168]. Im Gegensatz dazu zeigte eine Studie von Caminero *et al.*, dass extrem hohe Raten an Mischinfektionen in Regionen mit geringer TB-Inzidenz auftreten [22]. Die Frequenzen variieren zwischen 2,8 % und 19 % in Abhängigkeit der geographischen Lokalisation und der durchgeführten Studie [74]. Neben der TB-Inzidenz als treibender Faktor könnte ebenfalls die inakkurate Detektionsmöglichkeit als Erklärung für diese Variation herangezogen werden (Abbildung 1.4) [134].

Insgesamt werden neue, bessere molekularbiologische Methoden zur Detektion von Resistenzen benötigt, um eventuelle Heteroresistenzen frühzeitig zu erkennen und eine adäquate, individualisierte Therapie zu initiieren [80].

Ein wesentlicher Fortschritt stellt hierbei der Einsatz von genom-basierten Typisierungsmethoden dar.

1.8 Genomsequenzierung von MTBK-Stämmen

Die Verwendung von Sequenzierungsdaten zur Resistenzbestimmung von MTBK-Stämmen ist ein vielversprechender Ansatz [33]. Seit 1975 wurde sich auf die Sequenzierung (Sanger-Sequenzierung) bestimmter Genomregionen, so genannter „hot spots“, der resistenzvermittelnden Gene fokussiert [62]. Mit der Entwicklung neuer Sequenzierverfahren (engl.: Next-Generation Sequencing, NGS) kann

heutzutage schnell und zuverlässig das gesamte Genom der Pathogene analysiert werden, was ein entscheidender Vorteil gegenüber der phänotypischen Resistenzbestimmung ist [21]. Die Sanger-Sequenzierung gehört zu den traditionellen, molekulargenetischen Sequenzierungstechniken. Bei dieser Methode werden im Vergleich zum gesamten Genom nur kleine Regionen, beispielsweise ein Gen oder Genfragmente (bis circa 1000 Basenpaare [bp]), analysiert [62].

Moderne NGS-basierte Verfahren erlauben die parallele Sequenzierung ganzer Genome mehrerer klinischer Isolate. Bei der als Hochdurchsatzverfahren bezeichneten Methode wird das ganze Genom in einem „Shotgun“-Verfahren analysiert [184]. Es handelt sich dabei um eine ungerichtete Sequenzierung, die nicht auf bekannte Abschnitte des Genoms beschränkt ist.

Zurzeit werden diese Hochdurchsatzverfahren durch drei konkurrierende Technologien getragen:

- Illumina
- Ion Torrent
- Pac Bio RS

Eine ausführliche Übersicht der einzelnen Verfahren ist in Tabelle 1.2 zusammengefasst.

Das Verfahren der Firma Illumina steht seit 2006 zur Verfügung. Hierbei wird das Zielgenom mit Hilfe von Endonukleasen zufällig in Fragmente zerteilt („Shotgun“-Sequenzierung) und mit Adaptoren für die anschließende PCR versehen. Die Fragmente werden auf einem Glasträger immobilisiert und mit Hilfe einer „Brücken“-PCR (engl.: Bridge Amplification) amplifiziert. Im Anschluss werden die Fragmente sequenziert [76]. Das verwendete Verfahren wird als „sequencing-by-synthesis“-Verfahren bezeichnet, da nach jedem Syntheseschritt des DNS-Stranges analysiert werden kann, welches Nukleotid (Adenin [A], Cytosin [C], Guanin [G] oder Thymin [T]) eingebaut wurde. Ähnlich, wie bei der Sanger-Sequenzierung werden fluoreszierende Kettenabbruchnukleotide eingesetzt, deren Einbau von speziellen Optiken detektiert wird [184]. Die Leseweite liegt zurzeit zwischen 51 bp und 351 bp. Die Illumina Technologie bietet die Möglichkeit der so genannten „paired-end“-Sequenzierung [85]. Bei dieser Methode wird ein Fragment jeweils von beiden Seiten abgelesen. Auf diese Weise entstehen zwei NGS-Sequenzfragmente (51 bp-351 bp), die einen bekannten Abstand zueinander haben, wodurch die Ausrichtung zu einer Referenzsequenz erleichtert wird [10, 85]. Außerdem lassen sich größere

Indels detektieren, wenn bei perfekter Ausrichtung der NGS-Sequenzfragmente zur Referenzsequenz zum Beispiel der Abstand der zwei NGS-Sequenzfragmente nicht dem erwarteten Abstand entspricht. Die in dieser Arbeit verwendeten NGS-Daten sind hauptsächlich mit Hilfe von Illumina HiSeq 2500, MiSeq oder NextSeq 500 Geräten erhoben worden.

Das Verfahren von Ion Torrent basiert auf einer Halbleitertechnologie und wurde 2010 eingeführt [173]. Es werden ebenfalls DNS-Fragmente des Gesamtgenoms mit Hilfe von Endonukleasen erzeugt. Diese werden mit einer Emulsions-PCR (emPCR) vervielfältigt. Die Sequenzierung unterscheidet sich jedoch von der Illumina-Methode. Es werden zwar ebenfalls sequentiell nacheinander die verschiedenen Nukleotide zugegeben, allerdings wird bei diesem Verfahren die Veränderung des pH-Werts gemessen [173]. Beim Einbau eines Nukleotids entsteht eine kovalente Bindung unter Freisetzung eines Pyrophosphats und eines positiv geladenen Wasserstoffions, in dessen Folge der pH-Wert ansteigt. Eine pH-sensitive Schicht detektiert diese Veränderung und ein Detektor registriert, in Abhängigkeit der zugegebenen Base, den Einbau. Bei Homopolymeren steigt der pH-Wert proportional zur Anzahl der gleichen Basen. Die Leseweite liegt zwischen 200 bp und 300 bp [173].

Das PacBio RS System von Pacific Bioscience gehört zu den so genannten Sequenziergeräten der dritten Generation, hier werden Leseweiten von bis zu 15 000 bp erzielt (Tabelle 1.2). Dabei werden einzelne DNS-Moleküle während der Replikation durch die DNS-Polymerase sequenziert (engl.: Single-Molecule Real-Time; SMRT) [58, 76]. Eine vorherige Amplifikation ist nicht nötig [76]. Die vier Nukleotide werden mit einem Fluoreszenz-Farbstoff versehen, der nach Einbau des Nukleotids von der DNS-Polymerase abgetrennt wird. Das Fluoreszenzsignal eines Nukleotids bei Einbau wird detektiert und in die DNS-Sequenz übersetzt. Die Fehlerrate ist jedoch mit etwa 14% im Vergleich zu anderen Methoden sehr hoch [58].

Bei allen hier vorgestellten NGS-Verfahren ist die Genauigkeit „pro gelesene Base“ signifikant geringer als bei der klassischen Sanger-Sequenzierungsmethode. Dies wird jedoch durch die Mehrfachlesung der Sequenzen kompensiert [76]. Es entsteht eine überlappende Sequenzfragment-Population aus der das Zielgenom rekonstruiert werden kann. Die umfangreiche Analyse der von den NGS-Verfahren millionenfach generierten kurzen Sequenzfragmente kann mit verschiedenen Methoden erfolgen:

1. *De-novo* Assemblierung
2. Ausrichtung zu einer Referenzsequenz

Bei der *de-novo* Assemblierung werden die Sequenzfragmente anhand ihrer Überlappungen zu größeren so genannten „Contigs“ zusammengefasst [10]. Auf diese Weise wird das Zielgenom rekonstruiert und eine so genannte Konsensussequenz abgebildet. Bei der Ausrichtung zu einer Referenzsequenz wird für jedes Sequenzfragment die Position in einem bekannten Referenzgenom ermittelt, an der es die meisten Übereinstimmungen mit der Referenzsequenz hat [118]. Im Anschluss kann die Probe auf Unterschiede zur Referenzsequenz untersucht werden und auf diese Weise das Zielgenom rekonstruiert werden.

Diese Analysen erfordern umfangreiche, computergestützte Verfahren, die insbesondere durch Kommandozeilenanwendungen (Computerprogramme ohne grafische Benutzeroberfläche) geprägt sind (Kapitel 2.3). Für Anwender ohne informatischen Hintergrund sind diese nur bedingt zugänglich [56]. Ein weiteres Problem bei der Analyse von NGS-Datensätzen ist deren Größe, die sich zwischen 150 Megabyte (MB) und drei Gigabyte (GB) bewegt. Die benötigte Rechen- und Speicherkapazität erfordert unter anderem eine umfangreiche Informationstechnik-Infrastruktur (IT-Infrastruktur) [56].

Um diese Probleme zu adressieren, wurden bereits benutzerfreundliche Programme entwickelt, die insbesondere die Auswertung von NGS-Daten zur Resistenzbestimmung bei TB-Erregern automatisieren und erleichtern sollen [18, 33, 56, 86, 192]. Ein systematischer Vergleich der vorhandenen Analysemöglichkeiten zur Resistenzvorhersage wurde jedoch noch nicht durchgeführt.

Tabelle 1.2: Sequenzierungsmethoden der marktführenden Firmen Illumina, Life Technologies und Pacific Bioscience im Vergleich. Quellen: [73, 103, 104]

Firma	Plattform	Verriegelungs- schritt	Sequenzier- prinzip	Signal Detektions- methode	Häufigster Sequenzier- fehler	Sequenz- fragment- länge	Lauf- zeit in h	Fehler- rate in %
Illumina	HiSeq2000		reversibler Farbstoff- Terminatoren	Optische Detektion	Substi- tutionen, meist zum Ende des Sequenz- fragments	bis 250bp	7-60	0.1
	Illumina MiSeq	Brücken-PCR auf Durchflusszelle	Sequenzie- rung während Synthese	der Fluores- zenzmission nach Einbau markierter Nukleotide		50-300bp	4-39	0.1
Life Techno- logies/ Ion	Illumina NextSeq					150bp	11-29	0.1
	Ion PGM	Emulsions- PCR mit Sequenzier- beads	Halbleiter- basierte Sequenzie- rung während Synthese	Transistor- basierte Detektion des pH-Wert- Anstiegs nach Einbau eines Nukleotids	Insertionen/ Deletionen	200- 300 bp	3	1
Pacific Bioscience	PacBio	nicht verwendet	Einzel- Molekül Echtzeit DNS Sequenzie- rung während Synthese	Optische Detektion eines Fluores- zenzfarbstoffs während des Einbaus in Echtzeit	Insertionen/ Deletionen	1300 bp	0.5-6	14
	RS							

1.9 Zielsetzung

Das vermehrte Auftreten resistenter MTBK-Stämme erschwert die weltweite Kontrolle und stellt eine der größten Herausforderungen bei der zukünftigen Bekämpfung der TB dar [158]. Für eine erfolgreiche Kontrolle von M/XDR-Stämmen ist die schnelle Detektion vorhandener Antibiotikaresistenzen als Basis für eine effektive Therapie von besonderer Bedeutung [158]. Verfahren, die auf der klassischen Resistenztestung beruhen, sind zeitintensiv und nur bedingt für Zweitrangantibiotika durchführbar [102]. Auch molekulare Schnelltests zur Resistenztestung sind nur eingeschränkt für Zweitrangantibiotika anwendbar und bilden nur einen geringen Teil der Genomsequenz ab [5, 129, 193].

NGS-basierte Verfahren ermöglichen einen schnellen und zuverlässigen Nachweis nahezu aller mit Resistenz assoziierten Varianten in einem Genom (Resistom) [158]. Darüber hinaus können klinische Isolate präzise in verschiedene phylogenetische Linien eingeteilt und eine hochauflösende Genotypisierung für Ausbruchuntersuchungen durchgeführt werden [108].

Zurzeit ist die routinemäßige Anwendung von NGS-basierten Verfahren für diagnostische Fragestellungen, aufgrund der nötigen Infrastruktur und einer umfangreichen Datenanalyse, nur eingeschränkt möglich [215]. Außerdem sind die frei verfügbaren Analyseprogramme noch nicht hinreichend evaluiert worden.

Auch der NGS-basierte Nachweis heterogener Populationen in klinischen Isolaten, der insbesondere für die Resistenzdiagnostik von großer Bedeutung ist, ist noch nicht ausreichend untersucht und beschrieben worden [134].

Um diese Fragestellungen zu adressieren, soll in dieser Arbeit Folgendes gezeigt werden:

1. Systematischer Vergleich automatisierter NGS-Analyseprogramme:
 - a) Vergleich der Handhabung und Darstellung der Ausgabe fünf automatischer Analyseprogramme (CASTB, KvarQ, Mykrobe Predictor TB, PhyResSE und TBProfiler) zur Resistenzvorhersage von TB-Erregern
 - b) Evaluierung der Güte der Resistenzvorhersage der Programme im Vergleich zur Sanger-Sequenzierung (Goldstandard) anhand eines gut beschriebenen Datensatzes

- c) Evaluierung der Güte der Genotypklassifikation im Vergleich zu standardisierten Verfahren (MIRU-VNTR Typisierung, IS6110 DNS Fingerprint Typisierung, Spoligotypisierung)
 - d) Analyse der Möglichkeit der Detektion von Heteroresistenzen/Mischinfektionen
2. Implementierung und Anwendung eines NGS-Mischungsdetektors:
- a) Implementierung eines Mischungsdetektors anhand von NGS-Daten
 - b) Anwendung des Detektors bei unterschiedlichen Populationen
3. Bewertung von niederfrequenten Resistenzmutationen:
- a) Ansatz zur Bewertung niederfrequenter Mutationen
 - b) Untersuchung serieller Daten auf niederfrequente Subpopulationen

Insgesamt soll diese Arbeit einen Beitrag zur verbesserten Analyse von Hochdurchsatzdaten der TB-Erreger leisten. Insbesondere die Möglichkeit der Detektion von Heteroresistenzen soll fokussiert behandelt werden.

2 Grundlagen

In diesem Kapitel wird auf die für diese Arbeit relevanten Grundlagen der Genomsequenzierung sowie der damit verbundenen Analyse der Sequenzierungsdaten eingegangen.

2.1 Illumina-Sequenzierung

Die in dieser Arbeit verwendeten Daten basieren auf der Illumina-Sequenzierung. Für das Herstellen der so genannten „Libraries“ (DNS-Bibliotheken) wurde das „Nextera XT Preparation Kit“ [4] und das von Baym *et al.* [13] 2015 beschriebene Verfahren nach Herstellerangaben verwendet. Beide Protokolle basieren auf demselben Grundprinzip und lassen sich in fünf Schritte einteilen (Abbildung 2.1), allerdings ist das Baym-Protokoll die günstigere Alternative [4, 13].

2.2 Ausgabeformate

Bei der Prozessierung von NGS-Daten gibt es für die unterschiedlichen Analyseabschnitte definierte Ausgabeformate. Nur so ist es möglich, dass Analyseprogramme unterschiedlicher Entwickler kompatibel miteinander arbeiten können. Die NGS-Primärdaten liegen im so genannten FastQ-Format vor. Im ersten Schritt werden die NGS-Sequenzfragmente der FastQ-Datei zu einer Referenzsequenz ausgerichtet (engl.: Mapper). Das Ergebnis der Ausrichtung wird im so genannten Sequenz-Ausrichtungs-Format (engl.: Sequence Alignment Map Format, SAM) ausgegeben. Varianten beziehungsweise Mutationen werden durch den Vergleich mit der Referenzsequenz terminiert und ebenfalls in dem speziell dafür entwickelten Varianten-Format (engl.: Variant Call Format, VCF) abgelegt. Weiterhin wird ein Format mit Komma-getrennten Werten (engl.: Comma-separated

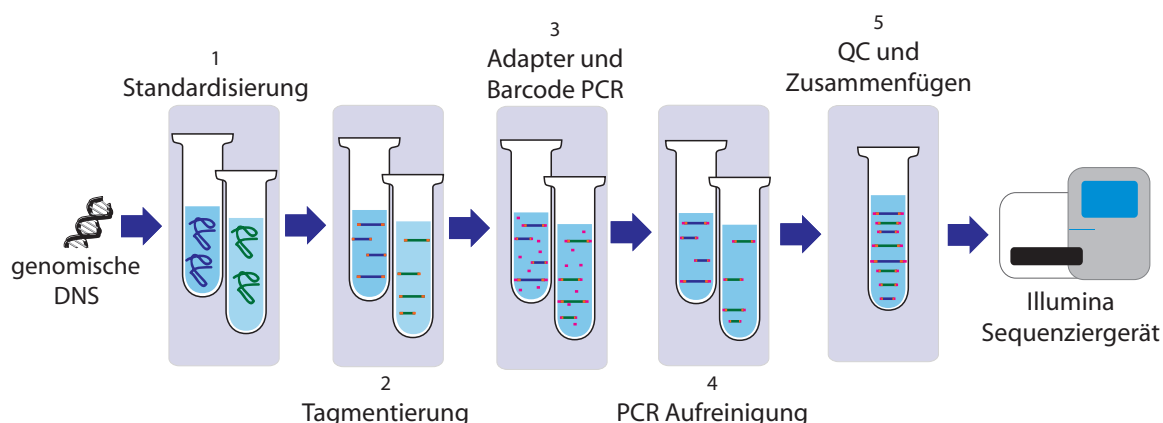


Abbildung 2.1: Schematische Darstellung der Illumina DNS-Bibliothekenerstellung beginnend mit der genomischen DNS bis zum Illumina-Sequenziergerät.

Standardisierung: Für eine einheitliche Reaktionseffizienz im nächsten Schritt (Tagmentierung) muss die Konzentration der genomischen DNS für alle Proben etwa gleich sein. Die optimale Konzentration liegt zwischen $0,5$ und $1 \text{ ng}/\mu\text{l}$ und kann unter anderem mit dem „QuBit“ gemessen werden.

Tagmentierung: Bei der Tagmentierung der genomischen DNS handelt es sich um eine modifizierte Transpositionsreaktion. Hierbei wird die genomische DNS gleichzeitig in Fragmente zerbrochen und mit Adaptoren für die Taq-Polymerase der PCR versehen. Dies geschieht zum Beispiel mit Hilfe des Nextera XT Transposons. Die Transposase schneidet die genomische DNS an zufälligen Stellen, sodass eine sich überlappende Molekülpopulation entsteht.

Adapter- und Barcode-PCR: Vervielfältigung der Fragmente mittels PCR und Anhängen von eindeutigen Identifikationssequenzen (Index-Sequenzen).

PCR Aufreinigung: Aufreinigung des PCR-Produktes mit Hilfe paramagnetischer Mikroperlen. Die Fragmente werden an die Mikroperlen gebunden und mit Hilfe einer magnetischen Platte separiert, sodass die Überstände abgenommen werden können. Danach schließen sich mehrere Waschgänge mit 80% Ethanol an.

Zusammenfügen der NGS-Bibliotheken: Zusammenfügen gleicher Volumina normalisierter NGS-Bibliotheken.

Beladen des Illumina-Sequenziergeräts.

values, CSV) und Daten im JavaScript-Objekt Format (JavaScript Object Notation, JSON) in dieser Arbeit verwendet, die eine Rolle als Ausgabeformate bestimmter NGS-Analyseprogramme spielen. In den folgenden Kapiteln werden die genannten Formate kurz beschrieben.

2.2.1 FastQ-Format

Die Primärdaten heutiger NGS-Geräte werden im FastQ-Format ausgegeben. Es unterscheidet sich von dem für Sequenzen verwendeten Fasta-Format durch zusätzliche Qualitätsinformation (Q) der determinierten Basen. Dabei wird jedes NGS-Sequenzfragment durch vier fest vorgeschriebenen Zeilen definiert (Quelltext 2.1).

Quelltext 2.1: Drei NGS-Sequenzfragmente im FastQ-Format.

```
@M00553:3:AOU76:1:1101:16419:1569 1:N:0:1
NTCGACACCTTGATCAAGAATCTTGGAACGCCAAGGAGGGCGCGAGTGCAAGAGCTTTCGGGAAGATCATGCCGACCAGGTGG
+
#555<<<<@@@9<@@ACCCCCEBEBAE+6>ECEDE+7+@>CD>>>55A-A-A+ACEEDEED@)+555=CCC=)4CA++<@@
M00553:3:AOU76:1:1101:15144:1570 1:N:0:1
NTACCAGAGCCAGGGCGCGATCGACCTGGCCTGCCGGCTACGTGAGCGTATCGGCGATCTCGACCAGATCGCGTCGATCGTGCTG
+
#5<???BBDDBDDDDGGGGGGHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
M00553:3:AOU76:1:1101:16096:1571 1:N:0:1
NGCGTAGCCTGCTTGCCTACCAATGGGCACACCCCGCAAGCAATTGCTGTTCATGGGTCAGGAATTCCGCCAACGCGCGCAATG
+
#5<???BADDEEDEDGGGGGIIIIIIHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
```

Zu Beginn eines jeden NGS-Sequenzfragments steht ein „@“ gefolgt von dem eindeutigen Identifikator für die Sequenz. Der Identifikator enthält unter anderem Informationen über das verwendete Sequenziergerät, den Sequenzierlauf und die verwendete Indexsequenz. Anschließend folgt eine zweite Zeile mit der Basenabfolge der Sequenz (Quelltext 2.1), gefolgt von einem „+“ in der dritten Zeile, in der der Identifikator obligat wiederholt auftritt. In der letzten Zeile ist die Qualitätsinformation (Q) der Basen aus der zweiten Zeile mit dem Amerikanischen Standardcode zum Informationsaustausch (engl.: American Standard Code for Information Interchange, ASCII) kodiert gegeben (Tabelle 2.1). Jeder Base wird genau ein Qualitätssymbol zugeordnet, sodass diese Zeile dieselbe Anzahl an Symbolen haben muss, wie die Zeile der Basenabfolge. Die Berechnung des Qualitätswerts Q anhand der Fehlerwahrscheinlichkeit e ist in der Gleichung 2.1

Tabelle 2.1: Qualitätswerte (Q) mit zugehörigen Fehlerwahrscheinlichkeiten (e) und korrespondierenden ASCII-Zeichen

Q	e	ASCII	Q	e	ASCII
0	1,0000000	!	22	0,00630957	7
1	0,7943282	"	23	0,00501187	8
2	0,6309573	#	24	0,00398107	9
3	0,5011872	\$	25	0,00316228	:
4	0,3981072	%	26	0,00251189	;
5	0,3162278	&	27	0,00199526	<
6	0,2511886	'	28	0,00158489	=
7	0,1995262	(29	0,00125893	>
8	0,1584893)	30	0,00100000	?
9	0,1258925	*	31	0,00079433	@
10	0,1000000	+	32	0,00063096	A
11	0,0794328	,	33	0,00050119	B
12	0,0630957	-	34	0,00039811	C
13	0,0501187	.	35	0,00031623	D
14	0,0398107	/	36	0,00025119	E
15	0,0316228	0	37	0,00019953	F
16	0,0251189	1	38	0,00015849	G
17	0,0199526	2	39	0,00012589	H
18	0,0158489	3	40	0,00010000	I
19	0,0125893	4	41	0,00007943	J
20	0,0100000	5	42	0,00006309	K
21	0,0079433	6			

dargestellt.

$$Q = -10 * \log_{10}(e) \quad (2.1)$$

Die Kodierung des Qualitätswerts Q mittels der ASCII-Zeichen ermöglicht die Darstellung aller auftretender Qualitätswerte mit nur einem Zeichen. Für die Kodierung wird zum berechneten Qualitätswert die Zahl 33 addiert, sodass die ASCII-Zeichen 33 (erstes darstellbares Zeichen) bis 126 auftreten können (Anhang A) [28]. Wie die Qualitätswerte Q mit der Fehlerwahrscheinlichkeit e und dem ASCII-Zeichen in der FastQ-Datei zusammenhängen, ist in Tabelle 2.1 dargestellt.

Durch die spezielle Darstellung über den negativen dekadischen Logarithmus korrelieren kleine Fehlerraten mit großen Qualitätswerten und kleine Qualitätswerte weisen auf eine unsichere Vorhersage der Base hin [53]. Basen mit einem Qualitätswert unter Q20 ($> 1\%$ Fehlerwahrscheinlichkeit) sind unzureichend, während Basen mit einem Qualitätswerte über Q28 ($< 0,2\%$ Fehlerwahrscheinlichkeit) als Basen mit sehr hoher Qualität angesehen werden [148]. Der Wert Q20 wird häufig als Grenzwert zur Verwendung einzelner Basen genutzt, da ab diesem Wert die Fehlerwahrscheinlichkeit vergleichsweise konstant bleibt (Abbildung 2.2) [148]

Eine FastQ-Datei enthält üblicherweise Millionen von NGS-Sequenzfragmenten für eine Probe, die bei der für diese Arbeit verwendeten Methode eine Länge von 151 bis 301 bp haben.

Durch die Verwendung der „paired-end“ Sequenzierung werden jeweils zwei FastQ-Dateien pro NGS-Probe ausgegeben.

2.2.2 Sequenz-Ausrichtungs-Format (SAM)

Das SAM-Format enthält Informationen zur Ausrichtung der NGS-Sequenzfragmente zur Referenzsequenz. Die Datei im SAM-Format ist ein einfaches Textdokument, in der in jeder Zeile ein NGS-Sequenzfragment repräsentiert wird. Die SAM-Datei besteht aus dem Dateikopf mit einer allgemeinen Beschreibung (Zeilen beginnen mit einem @) und dem Dateikörper, der Informationen über die Ausrichtung der jeweiligen NGS-Sequenzfragmente zur Referenzsequenz beinhaltet. Jede Zeile des Dateikörpers besteht aus elf obligatorischen Spalten und beliebig vielen optionalen Spalten. Einer der wichtigsten Werte des SAM-Formats ist zum Beispiel die Qualität der Ausrichtung (engl.: Mapping Quality, MAPQ). Dieser

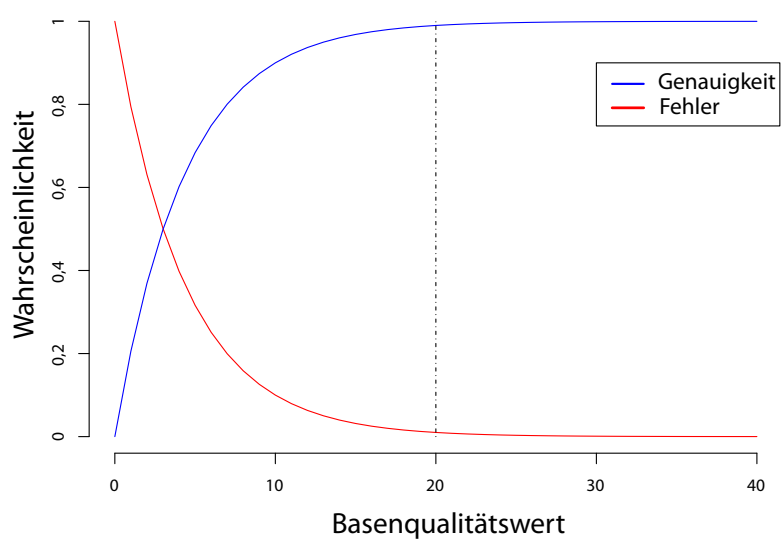


Abbildung 2.2: Darstellung der Fehlerwahrscheinlichkeit e (Genauigkeit) einer Base in Abhängigkeit des Basenqualitätswerts. Ab einem Qualitätswert von Q20 ($> 1\%$ Fehlerwahrscheinlichkeit) ist nur noch ein flacher Anstieg (Genauigkeit, blaue Linie) beziehungsweise Abfall (Fehler, rote Linie) des Kurvenverlaufs zu beobachten, sodass sich dieser Wert als Grenzwert etabliert hat. Erstellt mit dem Statistikprogramm R Version 3.1.2.

Wert gibt für jedes NGS-Sequenzfragment die Qualität der Übereinstimmung mit der ausgerichteten Position in der Referenzsequenz an; er berechnet sich mit

$$MAPQ = -10 \log_{10}(P(\text{ausgerichtete Position falsch}))$$

und wird auf die nächste ganze Zahl gerundet. Die exakte Spezifikation der obligatorischen Spalten ist in der „Sequence Alignment/Map Format Specification“ der SAM-Format Spezifikations-Arbeitsgruppe [130] gegeben. Die binäre Version des SAM-Formats heißt BAM-Format und enthält dieselbe Information [130].

2.2.3 Varianten-Format (VCF)

Das Varianten-Format (engl.: Variant Call Format, VCF) ist ein standardisiertes Format zum Abspeichern von DNS-Polymorphismen, wie SNPs und Indels [37]. Jede Zeile dieses Dokuments repräsentiert eine Mutation. Hinter den acht obligatorischen Spalten, in denen unter anderem Informationen über die Referenz- sowie die alternative Base einer variable Position abgelegt sind, kann optional eine Format-Spalte und beliebig viele Probenspalten auftreten. Jede Probenspalte erhält die Information zu einer NGS-Probe.

2.2.4 Komma-getrennte Werte-Format (CSV)

„CSV“ steht dabei für das englische „comma-separated values“ und besteht demnach aus Werten, die mittels Komma voneinander getrennt sind. Dieses Format wird auch „character-separated values“ übersetzt, was der korrekteren Bezeichnung entspricht, da nicht zwingend ein Komma als Feldtrennzeichen verwendet wird. Das CSV-Format beschreibt eine Textdatei zum Austausch strukturierter Daten zwischen unterschiedlichen Programmen. Es gibt vielmehr verschiedene Spezifikationen und Implementationen des CSV-Formats [46, 49, 172].

2.2.5 JavaScript-Objekt-Format (JSON)

Das JSON-Format ist ein Textformat für den einfachen Datenaustausch [48]. Das Format ist nicht von einer Programmiersprache abhängig und erfüllt Konventionen für C-basierte Programmiersprachen. Das Format baut auf zwei grundlegenden Strukturen auf:

- Name/Wert Paare
- Geordnete Listen von Werten

Die JSON-Datei enthält mehrere „Objekte“, die jeweils durch geschwungene Klammern abgegrenzt sind und eine ungeordnete Menge von Name/Wert Paaren, die jeweils durch ein Komma voneinander getrennt sind, enthält. Nach einem Namen steht ein Doppelpunkt gefolgt von dem Wert. Ein Wert im JSON-Format kann ein weiteres Objekt, ein Liste (engl.: Array), eine Zeichenkette (engl.: String), eine Zahl oder einer der Ausdrücke *true*, *false* oder *null* sein. Eine Zeichenkette wird von doppelten Anführungszeichen umschlossen.

Die geordnete Liste von Werten wird durch eckige Klammern begrenzt, wobei die Werte wiederum mit einem Komma voneinander getrennt sind (Minimalbeispiel in Quelltext 2.2).

Quelltext 2.2: Minimalbeispiel zweier Objekte im JSON-Format.

```
{}`Wertname1` : ``Wert1``, ``Wertname2`` : ``Wert2``, ``Listenname1`` : [``Wertname3`` :  
  ``Wert3`` , ``Wertname4`` : ``Wert4``]},  
{}`Wertname5`` : ``Wert5``, ``Wertname6`` : ``Wert6``}
```

2.3 Analyse von NGS-Daten

Zurzeit gibt es keine einheitliche Empfehlung zur standardisierten Analyse von NGS-Daten. Die Art der Generierung der NGS-Daten (Kapitel 1.8), impliziert jedoch obligatorische Analyseschritte [6], die im Folgenden kurz erläutert werden.

Basenbestimmung: Die Verfahren zur Bestimmung der Basensequenz einer biologischen Probe wurden bereits in Kapitel 1.8 dargestellt. Die Ausgabe erfolgt im FastQ-Format (Kapitel 2.2.1).

Qualitätskontrolle: Die unprozessierten Sequenzierungsdaten werden mit bestimmten Programmen (z. B.: FastQC) auf ihre Qualität geprüft. Zu den Qualitätswerten zählen unter anderem die Verteilung der Qualitätswerte und der GC-Gehalt der NGS-Sequenzfragmente. In diesem Schritt ist es möglich, NGS-Sequenzfragmente heraus zu filtern, die bestimmte Qualitätskriterien nicht erfüllen (z. B.: 90 % der Basen zeigen Qualitätswert > 20).

Ausrichtung/Assemblierung: Es gibt zwei grundsätzlich unterschiedliche Verfahren aus den NGS-Sequenzfragmenten das Zielgenom zu rekonstruieren (Kapitel 1.8). In dieser Arbeit wird die Ausrichtung zu einer Referenzsequenz verwendet. Für diese Aufgabe gibt es eine Vielzahl unterschiedlicher Programme, die Dateien im SAM- beziehungsweise BAM-Format ausgeben.

Post-Prozessierung: Je nach verwendetem Programm und Organismus sollten die initialen Rekonstruktionen des Zielgenoms nachgebessert werden. Zur Verbesserung der Ausrichtung der NGS-Sequenzfragmente zur Referenzsequenz kann beispielsweise eine Neu-Ausrichtung der NGS-Sequenzfragmente um kleine Indels oder das Entfernen von PCR-Artefakten (Duplizierte NGS-Sequenzfragmente) nötig werden. Das Genomanalyse Softwarepaket (engl.: Genome Analysis Toolkit; GATK) vom Broad Institut und SAMtools bieten für diese Verarbeitungsschritte eine Vielzahl Algorithmen an [42, 119, 135].

Qualitätswertrekalibrierung: Es wurde beschrieben, dass die Software der heutigen Sequenziergeräte die Basenqualität im Allgemeinen überschätzen [121]. Aus diesem Grund wird die Rekalibrierung der Basenqualitätswerte empfohlen [6], für die unter anderem GATK Algorithmen zur Verfügung stellt.

Variandendetektion: Nachdem das Zielgenom rekonstruiert wurde, sind im Allgemeinen die SNPs im Vergleich zur Referenzsequenz von Interesse. Die verfügbaren Programme geben die Ergebnisse im VCF-Format aus (Kapitel 2.2.3).

Filterung der SNP-Kandidaten: Das Filtern der detektierten Mutationen, zum Beispiel nach Qualität oder biologischer Relevanz, vermeidet falsch-positive Ergebnisse. Algorithmen überprüfen beispielsweise die Nähe zu Indels oder ob bei „paired-end“ sequenzierten NGS-Proben ein Richtungs-Bias zu beobachten ist, also der detektierte SNP beispielsweise nur auf den vorwärtssequenzierten NGS-Sequenzfragmenten detektiert wurde.

Interpretation der SNP-Kandidaten: Am Ende der Analyse steht die Interpretation der SNP-Kandidaten.

Die Analyse von NGS-Datensätzen besteht demnach aus einer komplexen Kombination verschiedener Programme. Der Großteil dieser Programme sind Kommandozeilenanwendung ohne grafische Benutzeroberfläche (engl.: Graphical User Interface; GUI) und benötigt auf Grund der Datenmenge eine hohe Speicherplatz- und Prozessorleistung.

Die in dieser Arbeit für den Hauptteil der verwendeten NGS-Daten ausgeführte Analysepipeline ist die Grundlage für die Webseite „Phylo Resistance Search Engine“ (PhyResSE) und ist in Abbildung 2.3 dargestellt.

2.4 Verwendete Datensätze

In dieser Arbeit wurden selbst erhobene NGS-Daten sowie NGS-Datensätze aus dem Europäischen Nukleotid Archiv (engl.: The European Nucleotide Archive; ENA; [51]) aus sechs Studien unterschiedlicher geographischer Regionen verwendet. Eine umfangreiche Beschreibung der Datensätze befindet sich im Anhang B. Tabelle 2.2 gibt einen allgemeinen kurzen Überblick.

Darüber hinaus wurden artifizielle Datensätze generiert, die Mischinfektionen und heteroresistenten Populationen mit eindeutigem Resistenzstatus simulieren. Dazu wurden DNS-Proben in einem festgelegten Verhältnis gemischt. Zur Herstellung von Mischungen mit Stämmen zweier unterschiedlicher phylogenetischer Linien wurde die DNS zweier antibiotikasensibler Stämme gemischt. Die DNS der Probe 49-02, eines Stammes der Linie Beijing (Linie 2.2.1, ENA-Zugriffsnummer: HG813240) und der Probe 7199-99, eines Stammes der Linie Haarlem (Linie 4.1.2, ENA-Zugriffsnummer: HE663067), wurde im Verhältnis 1:1, 7:3 und 9:1 gemischt. Anschließend wurden die NGS-Bibliotheken für die Sequenzierung hergestellt (Kapitel 2.1) und auf dem Illumina MiSeq Gerät mit einer 100-fachen Referenzgenomabdeckung (engl.: Coverage) sequenziert, das heißt im Mittel wird jede Position der Referenzsequenz von 100 NGS-Sequenzfragmente abgedeckt. Zur Erstellung von heteroresistenten Proben wurde die DNS des Laborstammes *M. tuberculosis* H37Rv ATCC 27294 mit der DNS zweier Selektionsklone gemischt. Für diese Arbeit wurden Selektionsklone mit den zwei häufigsten

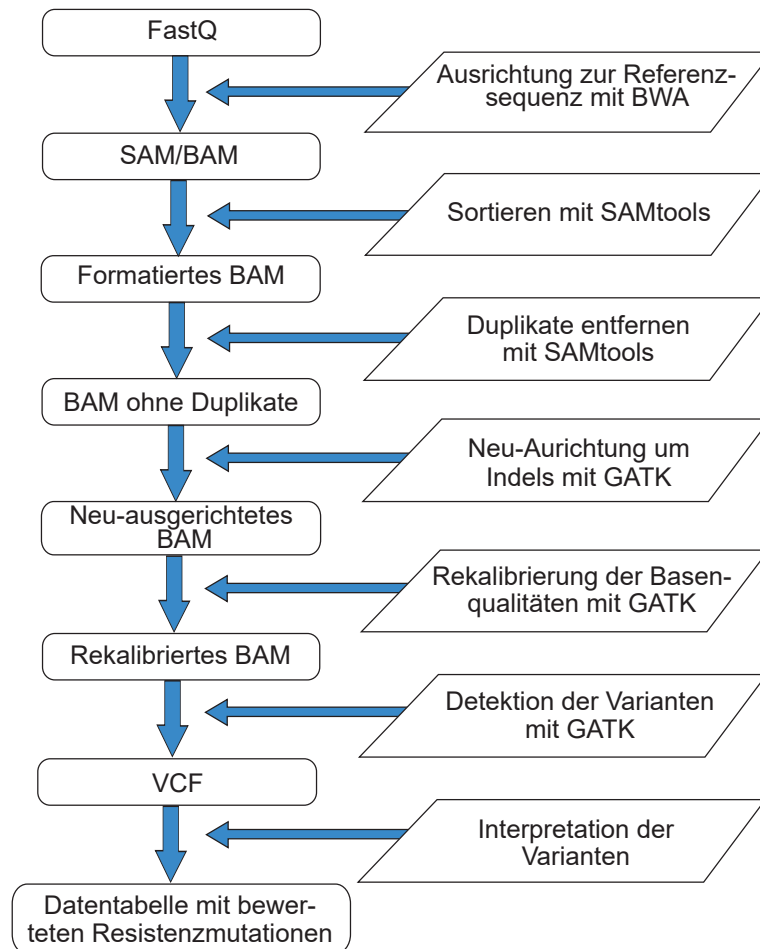


Abbildung 2.3: Abfolge der Analyseschritte von PhyResSE mit zugehörigen Programmen und resultierenden Formaten. Beginnend mit der FastQ-Datei wird mittels eines Algorithmus zur Ausrichtung von Sequenzfragmenten zu einer Referenzsequenz von BWA eine SAM-beziehungswise BAM-Datei erzeugt. Diese BAM-Datei wird mit SAMtools, einem Programm speziell zur Analyse von SAM-Dateien, weiter verarbeitet, sodass eine BAM-Datei erzeugt wird, die nach der ausgerichteten Position der Referenzsequenz sortiert und frei von duplizierten NGS-Sequenzfragmenten ist. In den darauffolgenden Schritten werden die NGS-Sequenzfragmente der BAM-Datei mittels Algorithmen von GATK um Indels neu zur Referenzsequenz ausgerichtet und die jeweiligen Basenqualitäten rekali­briert. Aus der resultierenden BAM-Datei wird ebenfalls mittels eines Algorithmus von GATK eine VCF-Datei erzeugt, in der die detektierten Mutationen im Vergleich zur Referenzsequenz abgebildet sind. Zum Schluss werden die detektierten Mutationen automatisch interpretiert und in Datentabellen ausgegeben.

Tabelle 2.2: Übersicht der verwendeten Datensätze (Stand 2015). Quellen: [43, 212]

Region	Allgemeine Daten					Genomdaten			
	Einwohnerzahl in Million	HDI ¹	TB-Inzidenz	MDR-TB ² in %	TB-Fälle ³ #	Anzahl #	Resistent #	MDR-Fälle #	EN-A-Nummer
Kenia	45	0,548	264	2,2	89 294	226	14	2	-
Russland	144	0,798	84	19	136168	1320	622	437	ERP000192
Sierra Leone	7	0,413	310	1,7	12 721	92	48	11	ERP008667
Swasiland	1,4	0,531	1382	7,7	7078	280	165	121	-
Turkmenistan	6,7	0,688	72	3,8	3675	71	64	51	-
Usbekistan	31,6	0,675	80	23	25 168	305	303	298	-

¹ Index der humanen Entwicklung (engl.: Human Development Index)

² Prozentualer Anteil von MDR-TB Fälle an allen TB-Neuerkrankungen

³ Gesamtanzahl TB-Erkrankungen (auch Rückfallpatienten und zuvor behandelte Patienten)

RMP-Resistenzmutationen *rpoB*531 (SR1a-Klon) und *rpoB*526 (SR4k-Klon) ausgewählt [143]. Es wurden jeweils Mischungen von Selektionsklon und Laborstamm im Verhältnis von 1:99, 1:19 und 1:9 generiert und mit einer mittleren Referenzgenomabdeckung von 500 NGS-Sequenzfragmenten pro Position auf dem Illumina NextSeq 500 Gerät sequenziert.

In Tabelle 2.3 sind die Bezeichnungen für die artifiziellen Datensätze, die im Folgenden verwendet werden, zusammengefasst.

Tabelle 2.3: Zusammenfassung der in dieser Arbeit verwendeten, artifizielle DNS-Proben

Bezeichnung	Beschreibung
Mix50	Artifizielle Mischung im Verhältnis 1:1 der Stämme 49-02 (ENA-Zugriffsnummern HG813240; Linie Beijing/Linie 2.2.1) und 7199-99 (ENA-Zugriffsnummer: HE663067; Linie Haarlem/Linie 4.1.2)
Mix70	Artifizielle Mischung im Verhältnis 7:3 der Stämme 49-02 (ENA-Zugriffsnummern HG813240; Linie Beijing/Linie 2.2.1) und 7199-99 (ENA-Zugriffsnummer: HE663067; Linie Haarlem/Linie 4.1.2)
Mix90	Artifizielle Mischung im Verhältnis 9:1 der Stämme 49-02 (ENA-Zugriffsnummern HG813240; Linie Beijing/Linie 2.2.1) und 7199-99 (ENA-Zugriffsnummer: HE663067; Linie Haarlem/Linie 4.1.2)
SR1a_rpoB531_1	Artifizielle Mischung eines Selektionsklons mit dem SNP <i>rpoB</i> S531L und dem Referenzstamm <i>M. tuberculosis</i> H37Rv ATCC 27294 im Verhältnis 1:99
SR1a_rpoB531_5	Artifizielle Mischung eines Selektionsklons mit dem SNP <i>rpoB</i> S531L und dem Referenzstamm <i>M. tuberculosis</i> H37Rv ATCC 27294 im Verhältnis 1:19
SR1a_rpoB531_10	Artifizielle Mischung eines Selektionsklons mit dem SNP <i>rpoB</i> S531L und dem Referenzstamm <i>M. tuberculosis</i> H37Rv ATCC 27294 im Verhältnis 1:9
SR4k_rpoB526_1	Artifizielle Mischung eines Selektionsklons mit dem SNP <i>rpoB</i> H526Y und dem Referenzstamm <i>M. tuberculosis</i> H37Rv ATCC 27294 im Verhältnis 1:99
SR4k_rpoB526_5	Artifizielle Mischung eines Selektionsklons mit dem SNP <i>rpoB</i> H526Y und dem Referenzstamm <i>M. tuberculosis</i> H37Rv ATCC 27294 im Verhältnis 1:19
SR4k_rpoB526_10	Artifizielle Mischung eines Selektionsklons mit dem SNP <i>rpoB</i> H526Y und dem Referenzstamm <i>M. tuberculosis</i> H37Rv ATCC 27294 im Verhältnis 1:9

2.5 Verwendete Software und Programmiersprachen

Die in der vorliegenden Arbeit erstellten Programme wurden größtenteils mit der Programmiersprache Perl implementiert. Zusätzlich wurden R-Pakete und der Bam-readcount Algorithmus genutzt. In den folgenden Abschnitten werden diese kurz beschrieben und erklärt.

2.5.1 Statistikprogramm R

Das Programm R bietet primär Möglichkeiten zur statistischen Datenanalyse. Es handelt sich um eine Kombination aus einer objekt-orientierten, interaktiven Programmiersprache und einer Laufzeitumgebung. R bietet eine flexible Grafikumgebung für die explorative Datenanalyse und eine Vielzahl klassischer sowie moderner statistischer und numerischer Verfahren. Dem Anwender stehen zahlreiche Ergänzungsmöglichkeiten zur Verfügung. Die Verwendung ist kostenfrei und der Quellcode frei erhältlich (engl.: Open Source Software). Für diese Arbeit wurde die R Version 3.0.2 verwendet.

2.5.2 Bam-readcount Algorithmus

Der Bam-readcount Algorithmus berechnet basierend auf einer BAM-Datei metrische Werte für einzelne Genompositionen. Die wichtigsten, hier verwendeten Werte sind:

base Base A, C, G oder T,

count beschreibt die Anzahl an NGS-Sequenzfragmenten, die diese Base zeigen

avg.basequality definiert die mittlere Basenqualität dieser Base

Die Benutzereingabe besteht aus der zu untersuchenden BAM-Datei und einer Liste der zu untersuchenden Genompositionen. Die Liste muss in der Form „Chromosom Start Stop“ für jeden Bereich angegeben werden, der analysiert werden soll. Ein Beispiel ist im Anhang C gegeben.

2.5.3 Programmiersprache Perl

Der Name Perl steht für „Practical Extraction and Report Language“. Perl ist eine freie, plattformübergreifende Programmiersprache. Ein Kompilieren, also in Maschinensprache umwandeln, ist nicht notwendig, da es sich um eine interpretierte Sprache handelt. Ursprünglich entstand Perl aus einer Kombination der Programmiersprache C und einigen UNIX-Funktionen. Perl ist nach der allgemeine Veröffentlichungsgenehmigung „GNU“ (engl.: GNU General Public License, GNU-GPL) einschließlich des Quelltextes frei verfügbar. In dieser Arbeit wurde die Perl Version 5.20.1 verwendet.

Teil II

Ergebnisteil

3 Automatisierte NGS-Analyseprogramme

Die Auswertung von NGS-Daten erfordert bioinformatische Kenntnisse, leistungsstarke Computer Hardware und Expertenwissen für die Interpretation der Daten und stellt somit eine Herausforderung für nicht spezialisierte Anwender zum Beispiel Biologen oder Kliniker dar. Außerdem erfordert jede biologische Fragestellung eine individuell angepasste Analyse. Für die NGS-basierte Resistenzdetektion von MTBK-Stämmen wurden beispielsweise bereits automatisierte Programme zur Analyse und Interpretation implementiert und veröffentlicht. Ein Vergleich der Programme im Hinblick zum Beispiel auf Anwenderfreundlichkeit und Sensitivität/Spezifität für die Vorhersage von Resistenzen aus NGS-Daten ist allerdings noch nicht durchgeführt worden. Die Programme werden im Folgenden kurz dargestellt und nachfolgend mit einem Testdatensatz evaluiert.

3.1 CASTB (Version 1.1)

CASTB steht für „The Comprehensive Analysis Server for the *Mycobacterium tuberculosis* complex“ und ist eine Webseite (<http://castb.ri.ncgm.go.jp/CASTB/>) [86], die zur Resistenzvorhersage und phylogenetischen Analyse von NGS-Daten klinischer MTBK-Stämme öffentlich zugänglich ist. Bevor jedoch Analysen durchgeführt werden können, ist eine Registrierung erforderlich. In der Eingabemaske können unter anderem FastQ-Dateien hochgeladen werden; ebenso ist es möglich assemblierte Genome oder assemblierte Genomabschnitte mit CASTB zu analysieren. Zur phylogenetischen Klassifizierung der MTBK-Stämme kann eine oder mehrere der folgenden Typisierungsmethoden ausgewählt werden, die im Anschluss virtuell basierend auf den NGS-Daten der hochgeladenen Probe durchgeführt werden:

- Spoligotypisierung (43 Spacer Sequenzen)
- Typisierung mittels langer Sequenz-Polymorphismen (LSP-Typisierung)
- Beijing-Subtypisierung
- MIRU-VNTR Typisierung (bei NGS-Sequenzfragmentlängen über 1000 bp möglich; 24 Loci)

Die Resistenzvorhersage wird für sechs Antituberkulotika (Tabelle 4.1) erstellt. CASTB liegt eine umfangreiche bakterielle Genomsequenz Datenbank (engl.: Bacterial Isolate Genome Sequence database; BIGSdb) zu Grunde [93]. Zum Assemblieren der NGS-Daten wird ein Algorithmenpaket namens „Velvet“ verwendet [218]. Neben einem Assembler ist ebenfalls der Algorithmus „MUMmer“ zum Ausrichten von NGS-Sequenzfragmenten zur Referenzsequenz *M. tuberculosis* H37Rv inkludiert [40,41,116]. Bei der phylogenetischen Klassifizierung der MTBK-Stämme wird, abhängig von der ausgewählten Methode, eine Klassifizierung in die sechs Hauptlinien des MTBK (nicht Linie 7/ Äthiopien) vorgenommen. Bei der Beijing-Subtypisierung werden die NGS-Sequenzfragmente einer Probe zu drei Zielsequenzen (non_beijin_1, beijing_1, beijin_2) mit „BLAST“ (engl.: Basic Local Alignment Search Tool) ausgerichtet, um zwischen Stämmen einer nicht-Beijing Linie und Stämmen der Linie Beijing zu unterscheiden. Zusätzlich werden die NGS-Sequenzfragmente zu den Sequenzen „ancestral_1“ und „modern_1“ geblas-tet, um Stämme der ancestralen und modernen Beijing Sublinie zu unterscheiden [137].

Die Resistenzvorhersage basiert laut Autoren von CASTB auf den resistenzvermittelnden Mutationen der Studien von Sekiguchi *et al.* und Ando *et al.* [7,8,181]; eine Liste der verwendeten resistenzvermittelnden Mutationen ist allerdings nicht verfügbar.

Die Ergebnisse der Analysen von CASTB werden in Form von Tabellen ausgegeben, die im CSV-Format heruntergeladen werden können. Die NGS-Daten, inklusive der Auswertungstabellen, werden sieben Tage auf dem CASTB-Server gespeichert.

3.2 KvarQ (Version 0.12.2)

KvarQ ist ein Computerprogramm zur direkten Variantendetektion aus FastQ-Dateien von MTBK-Stämmen [192]. Das Programm kann unter www.swisstph.ch/kvarq heruntergeladen werden. Es gibt neben einer Kommandozeilenanwendung, Implementationen für Windows und Mac OS X, die über eine grafische Benutzeroberfläche verfügen.

Das Programm untersucht für bestimmte Zielregionen der Genomsequenz des Referenzstamms *M. tuberculosis* H37Rv, die in Python-Modulen festgelegt sind, nach passenden NGS-Sequenzfragmenten in der FastQ-Datei und richtet diese entsprechend aus. Auf diese Weise wird auf den Schritt der Referenzgenomausrichtung verzichtet. In dieser Arbeit wurden die Standardparameter und die vorinstallierten Python-Module „MTBC/phylo“, „MTBC/resistance“ und „MTBC/s-poligo“ verwendet [192].

Die Stämme einer analysierten Probe werden in eine der sieben Hauptlinien des MTBK und die Sublinie Beijing klassifiziert. In dem Python-Modul „MTBC/phylo“ werden jeweils drei SNP-Positionen pro phylogenetischer Linie analysiert, auf denen die phylogenetische Klassifizierung der Stämme basiert. Zusätzlich kann eine virtuelle Spoligotypisierung durchgeführt werden.

Die Resistenzvorhersage wird für neun Antibiotika durchgeführt (Tabelle 4.1). Das Python-Modul „MTBC/resistance“ definiert insgesamt 561 bp des *pncA*-Gens, die 81 bp großen RMP-Resistenz assoziierte Kernregion (engl.: Rifampicin Resistance Determining Region; RRDR), die 63 bp großen Quinolon-Resistenz assoziierte Kernregion (engl.: Quinolone Resistance Determining Region; QRDR), 3 bp in *katG* und 32 SNP-Positionen in weiteren, bekannten resistenzvermittelnden Genen der Referenzsequenz *M. tuberculosis* H37Rv, die auf Mutationen untersucht werden (detaillierte Mutationsliste im Anhang D). Die Mutationsliste basiert hauptsächlich auf der Tuberkulose Resistenz-Datenbank (engl.: Tuberculosis Drug Resistance Mutation Database, TBDRaMDB) [175].

Die Analyseergebnisse von KvarQ können im JSON-Format abgespeichert und mit Hilfe der Grafikoberfläche untersucht werden.

3.3 Mykrobe Predictor TB (Version 0.1.3)

Mykrobe Predictor TB ist ein Programm zur schnellen Bestimmung von Resistenzmutationen in NGS-Daten von MTBK-Stämmen [18]. Das Programm kann unter <http://www.mykrobe.com/products/predictor/> heruntergeladen werden. Es gibt neben einer Kommandozeilenanwendung, Implementierungen für Windows und Mac OS X, die eine grafische Benutzeroberfläche haben.

Die Analyse-Pipeline besteht aus der Ausrichtung der NGS-Sequenzfragmente zur Referenzsequenz von *M. tuberculosis* H37Rv (GenBank NC000962.2) mit dem Ausrichtungsalgorithmus „Stampy“ [124] und der Variantendetektion mit Algorithmen von SAMtools/VCFtools [119].

Für die phylogenetische Klassifizierung werden Marker für vier MTBK-Spezies (*M. tuberculosis*, *M. africanum*, *M. bovis* und *M. caprae*) und 39 NTM (z. B.: *M. abscessus*, *M. aromaticivorans*, *M. avium*, u. A. siehe [18]) verwendet.

Für die Resistenzvorhersage werden Resistenzmarker für acht Antibiotika untersucht (Tabelle 4.1). Die Resistenzvorhersage von Mykrobe Predictor TB basiert auf den Mutationen, die für die molekularen Testsysteme von HAIN [27], Cepheid [129] und AID [170] verwendet werden sowie weiteren Mutationen aus der Literatur (detaillierte Liste im Anhang D).

Die Analyseergebnisse von Mykrobe Predictor TB können im JSON-Format abgespeichert und mit der grafischen Benutzeroberfläche erneut geöffnet und untersucht werden.

3.4 PhyResSE (Version 1.0)

Im Rahmen der Masterarbeit „Analyse von Next-Generation-Sequencing Daten zur Resistenzdetektion bei Tuberkulosebakterien“ [178] wurden die Grundsteine für den Webservice PhyResSE gelegt.

PhyResSE steht für „Phylo-Resistance Search Engine“ und ist eine Webseite zur automatisierten Analyse der NGS-Daten von MTBK-Stämmen zur phylogenetischen Klassifizierung und Resistenzvorhersage [56]. Die Seite ist über die Adresse www.phyresse.org erreichbar, eine Registrierung ist nicht erforderlich.

Die Analyse-Pipeline besteht aus der Ausrichtung der NGS-Sequenzfragmente zur

Referenzsequenz *M. tuberculosis* H37Rv (GenBank NC000962.3) mit dem Ausrichtungsalgorithmus „bwa mem“ von BWA [118], der Entfernung von Duplikaten (PCR-Artefakte) mit dem Algorithmus „samtools rmdup“ von SAMtools [119], der Neu-Ausrichtung der NGS-Sequenzfragmente um kleine Indels mit den Algorithmen „RealignerTargetCreator“ und „IndelRealigner“ von GATK, der Rekali-brierung der Basenqualitätswerte mit dem Algorithmus „BaseRecalibrator“ von GATK und der Variantendetektion mit dem Algorithmus „UnifiedGenotyper“ von GATK (Abbildung 2.3). Die phylogenetische Klassifizierung basiert auf spezifischen SNPs aus den Studien von Coll *et al.* 2014 [32] und Homolka *et al.* 2012 [82]. Das Resistenzprofil besteht aus der Resistenzvorhersage für zwölf Antibiotika (Tabelle 4.1) und basiert auf resistenzvermittelnden Mutationen verschiedener Studien einer umfangreichen Literaturrecherche (komplette Liste im Anhang D). Es ist möglich, die Ergebnisse einzelner NGS-Proben oder ganzer NGS-Studien herunterzuladen und im CSV-Format abzuspeichern (Kapitel 2.2.4). Die FastQ-Dateien sowie die Analyseergebnisse, werden über einen 32-stelligen Zahlencode, mit dem der Nutzer selektiv auf die persönlichen Daten zugreifen kann, gesichert und 3 Monate auf dem Server gespeichert.

3.5 TBProfiler

TBProfiler ist eine Webseite zur schnellen Detektion von Antibiotikaresistenzen in NGS-Daten von MTBK-Stämmen [33]. Die Webseite kann über die Adresse <http://tbdr.lshtm.ac.uk/> aufgerufen werden, eine Registrierung ist nicht erforderlich. Es ist ebenfalls möglich den Programmcode herunterzuladen und die Analysen von TBProfiler lokal über die Kommandozeile durchzuführen. In der Eingabemaske der Webseite können FastQ-Dateien von Illumina Sequenziergeräten hochgeladen werden. Die Analyse-Pipeline besteht aus der Ausrichtung der NGS-Sequenzfragmente zu einer modifizierte Version der Referenzsequenz *M. tuberculosis* H37Rv (spezielle Bereiche aus GenBank NC000962.3) mit dem Ausrichtungsalgorithmus „Snap“ [217] und der Variantendetektion mit SAMtools [119].

Die Ausgabe beinhaltet die Resistenzvorhersage für elf Antibiotika (Tabelle 4.1) sowie die phylogenetischen Klassifizierung anhand spezifischer SNPs. Die Muta-

tionsliste umfasst 1325 SNPs und Indels (detaillierte Liste im Anhang [D](#)). Die Ergebnisse werden unter einer Identifikationsnummer für alle Benutzer öffentlich gespeichert und können beliebig oft aufgerufen werden.

4 Evaluierung automatischer NGS-Analyseprogramme mit Hilfe eines Validierungsdatensatzes

Eine vergleichende Evaluierung der frei-verfügbaren Programme, zum Beispiel im Hinblick auf die Resistenzvorhersage, wurde bislang noch nicht durchgeführt. Im Rahmen der vorliegenden Arbeit wurden daher die NGS-basierte Resistenzvorhersage und die phylogenetische Klassifizierung von MTBK-Stämmen mit den frei-verfügbaren Programmen CASTB, KvarQ, Mykrobe Predictor TB, PhyResSE und TBProfiler mit einem gut charakterisierten NGS-Datensatz evaluiert [179]. Dazu wurden jeweils NGS-Daten von 91 Proben aus Sierra Leone (Kapitel 2.4) mit den Programmen zur automatisierten NGS-Analyse prozessiert. Dieser Datensatz wurde sehr gut charakterisiert [55, 81], indem zum einen die Stämme mittels der klassischen Typisierungsmethoden phylogenetisch klassifiziert wurden (Kapitel 1.5) und zum anderen Sanger-Sequenzen der Resistenzgene *ahpC*, *embA*, *embB*, *embC*, *gidB*, *inhA*, *katG*, *rrs*, *rpsL*, *pncA* und *rpoB* erstellt wurden. Die Ergebnisse der Analysen mit den verschiedenen Programmen wurden jeweils zusammengefasst und anschließend bezüglich der phylogenetischen Klassifizierung und der Resistenzvorhersage evaluiert.

4.1 Allgemeiner Vergleich

Die untersuchten Programme unterscheiden sich nicht nur bei den Analyseergebnissen, sondern auch bei der Bedienung, bei der Anpassbarkeit und weiterer Funktionen (Tabelle 4.2).

Ein wichtiger Aspekt hierbei ist die Anzahl der Antibiotikaresistenzen, auf die eine NGS-Probe untersucht werden kann. Diese variierte stark zwischen den evaluierten Programmen. Eine Übersicht ist in Tabelle 4.1 gezeigt. Für die Erstran-

gantibiotika sind außerdem die Gene dargestellt, in denen Mutationen für die Resistenzvorhersage detektiert werden.

Wie in Tabelle 4.1 zu erkennen ist, erlauben alle Programme eine Resistenzvorhersage für die Erstrangantibiotika INH, RMP und EMB. Mykrobe Predictor TB war das einzige Programm im Test, das keine Vorhersage für das Erstrangantibiotikum PZA erstellte. Insgesamt erlaubten PhyResSE und TBProfiler die umfassendste Analytik mit einer Resistenzvorhersage für bis zu zwölf Antibiotika pro NGS-Probe. CASTB hatte das kleinste Spektrum und erlaubte unter den Zweitrangantibiotika ausschließlich eine Resistenzvorhersage für CFX und SM (Tabelle 4.1).

Die Qualität der jeweiligen Resistenzvorhersagen der verschiedenen Programme für die Antibiotika INH, RMP, EMB, SM und PZA wird in Kapitel 4.3 behandelt.

Die Programme unterschieden sich außerdem in ihrem Layout und der Bedienung. In Tabelle 4.2 sind die wichtigsten Unterscheidungsmerkmale aufgeführt. Der größte Unterschied in Bezug auf die Handhabung der getesteten Programme bestand in der Verarbeitung der FastQ-Dateien. Für zwei Programme (KvarQ und PhyResSE) war beispielsweise eine „Batch“-Prozessierung der 91 MTBK-Stämme aus Sierra Leone möglich, während die NGS-Proben bei den anderen Programmen einzeln prozessiert werden mussten. Außerdem mussten die FastQ-Dateien der „paired-end“ sequenzierten Proben des Sierra Leone Datensatzes für die Verwendung mit KvarQ und MykrobePredictor TB im Vorfeld zusammengefügt werden, da nur eine Datei pro NGS-Probe analysiert werden konnte (Tabelle 4.2).

Ein weiteres Unterscheidungsmerkmal stellt die verwendete Analyse-Pipeline der getesteten Programme dar. Jedes Programm analysierte die FastQ-Dateien des Sierra Leone Datensatzes mit einer individuellen Analyse-Pipeline (Tabelle 4.2). KvarQ war in diesem Test das einzige Programm, das den zeitaufwendigen Schritt der Ausrichtung der NGS-Sequenzfragmente zu einer Referenzsequenz überspringt und mit einem k-mer Ansatz arbeitet. Auf diese Weise ist es theoretisch möglich eine NGS-Probe mit einer mittleren Referenzgenomabdeckung von 100 NGS-Sequenzfragmenten pro Base innerhalb von nur zwei Minuten mit KvarQ zu analysieren [192]. Im Gegensatz dazu beinhaltete die Analysepipeline von PhyResSE mehrere, rechenzeitaufwendige Vor-Prozessierungsschritte (Tabelle 4.2).

Ein weiterer Unterschied war bei der Darstellung und Exportierbarkeit der Ana-

Tabelle 4.1: Antibiotika beziehungsweise die untersuchten Resistenzgene für die von den getesteten Programmen eine Resistenzvorhersage erstellt wird. Modifiziert aus: [179]

Programm	Erstrangantibiotika										Zweitrangantibiotika									
	<i>ahpC</i>	<i>inhA</i>	<i>katG</i>	<i>ndh</i>	RMP	PZA	<i>emdB</i>	<i>emdB</i>	<i>emdB</i>	SM	AMK	CPM	KAN	ETH	FQ	CFX ¹	OFX ¹	MOX ¹	LZD	PAS
CASTB	?	✓	?	?	✓	?	✓	?	?	?	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Kvar-Q	✓	✓	✓	?	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mykr.Pred. TB	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
PhyResSE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
TBProfiler	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

✓ → Resistenzstatus für Antibiotikum/Gen wird ausgegeben; ? → keine Information verfügbar. Abkürzungen: Amikacin-AMK, Capreomycin-CPM, Ciprofloxacin-CFX, Ethambutol-EMB, Ethionamid-ETH, Fluorchinolon-FQ, Isoniazid-INH, Kanamycin-KAN, Linezolid-LZD, Moxifloxacin-MOX, Ofloxacin-OFX, Paraaminosalicylsäure-PAS, Pyrazinamid-PZA, Rifampicin-RMP, Streptomycin-SM.

¹ Gehört zur Gruppe der FQ.

² Indirekter Resistenzstatus durch FQ-Resistenzvorhersage.

4 Evaluierung automatischer NGS-Analyseprogramme mit Hilfe eines Validierungsdatensatzes

Tabelle 4.2: Funktionalität der fünf Programme zur automatisierten Analyse von NGS-Daten zur Resistenzvorhersage und phylogenetischen Klassifizierung von MTBK-Stämmen.

Funktion	CASTB (Version 1.1)	KvarQ (Version 0.12.2)	Mykrobe Predictor (Version 0.1.3)	TB	PhyResSE (Version 1.0)	TBProfiler
Web-basiert	Ja, mit Registrierung	Nein	Nein	Ja	Ja	Ja
Batch-Prozessierung	Nein	Ja	In der Kommandozeilen-Version	Ja	Nein	Nein
Paired-end sequenzierte Fragmente	Ja	In einer Datei	In einer Datei	Ja	Ja	Ja
Analyse-Pipeline	Velvet <i>de-novo</i> Assemblierung; BGSdb ¹ ; MUMmer Genomausrichtung; individuelle Skripte	Python-Module/Pakete mit C Erweiterungen (keine Referenzgenom-Ausrichtung)	Stampy Ausrichtung (H37Rv Version 2); Varianten-detektion SAMtools	BWA ² Referenzgenom-Ausrichtung (H37Rv Version 3); Neu-Ausrichtung, Rekalibrierung, Entfernung von Duplikaten GATK ³ Variantendetektion GATK ³	Referenzgenom-snap Ausrichtung (Teilregionen von H37Rv Version 3); Varianten-detektion SAMtools	Referenzgenom-Ausrichtung (Teilregionen von H37Rv Version 3); Varianten-detektion SAMtools
Textausgabe	Ja	Nein	Ja	Ja, ausführlich	Ja	Ja
Datenelexport	Ja	Ja	Nein	Ja	Nein	Nein
Detektion von NTM ⁴	Ja	Nein (nur MTBK)	Ja	Indirekt beim Datentransfer	Nein (nur MTBK)	Nein (nur MTBK)
Mutationsreport	Nein	Resistenzvermittelnde	Resistenzvermittelnde	Ja, alle	In Kandidatenregionen	In Kandidatenregionen
Justierbarkeit	Nein	Ja	Ja	Ja (undokumentiert)	Ja	Ja
Zusatz-funktionen	Spoligotypisierung, RFLP ⁵ LSP ⁶ , MIRU-VNTR ⁷ , Phylogenetischer Baum	Spoligotypisierung		SNP-Detektion, Phylogenetischer Baum		

¹ Bakterienproben Genomsequenz Datenbank (engl.: Bacterial Isolate Genome Sequence database)
² Ausrichtung mit der Burrows-Wheeler Transformation (engl.: Burrows-Wheeler Algorithm)
³ Genomanalyse Algorithmenansammlung (engl.: Genome Analysis Toolkit)
⁴ Nichtübertragbaren Mykobakterien
⁵ Restriktions-Fragmentlängen Polymorphismus
⁶ Großer Sequenzpolymorphismus (engl.: Large Sequence Polymorphism)
⁷ Mycobacterial Interspersed Repetitive Units-Variable Number of Tandem Repeats

lyseergebnisse der verschiedenen Programme zu beobachten. Während die Programme KvarQ und Mykrobe Predictor TB die Ausgaben nur in dem speziell auf die jeweilige grafische Benutzeroberfläche spezifizierten JSON-Format zur Verfügung stellten, konnten die Analyseergebnisse von CASTB und PhyResSE als CSV-Dateien exportiert werden. TBProfiler stellt alle analysierten Daten auf der Webseite zur Verfügung und liefert keine Lösung zum Exportieren der Analyseergebnisse.

Die Darstellung der Analyseergebnisse der getesteten Programme hat Einfluss auf die Anwenderfreundlichkeit der einzelnen Programme, da diese durch die gezielte Verwendung von Farben, Symbolen und unterschiedlichen Schriftarten erhöht wird [166]. Die Darstellung der Analyseergebnisse von Mykrobe Predictor TB und PhyResSE waren demnach besonders anwenderfreundlich, während die Ausgabe von KvarQ beispielsweise an die Ausgabe einer Kommandozeile erinnerte.

Doch nicht nur die optische Präsentation auch der Informationsgehalt der Programmausgaben unterschied sich voneinander (Tabelle 4.2). In der CASTB-Tabelle zur Resistenzvorhersage zeigte beispielsweise ein „R“ das Vorhandensein einer resistenzvermittelnden Mutation für das jeweilige Antibiotikum an; die detektierte resistenzvermittelnde Mutation wurde allerdings nicht ausgegeben. Die anderen Programme gaben mindestens den Aminosäureaustausch der resistenzvermittelnden Mutation an, auf der die Resistenzvorhersage basierte. Zusätzlich gaben PhyResSE und TBProfiler detektierte Mutationen in bekannten resistenzvermittelnden Genen, beziehungsweise alle detektierten Mutation im Genom (PhyResSE), aus.

Alle getesteten Programme sind im derzeitigen Entwicklungsstatus für die Verwendung in der Forschung bestimmt. Aus diesem Grund war die Justierbarkeit der Mutationslisten der verschiedenen Programme von Interesse, um zum Beispiel neue Kandidatenmutationen in einer Referenzkollektion auf ihre resistenzvermittelnde Funktion untersuchen zu können. CASTB bot keine Möglichkeit diese Problemstellung zu adressieren. Bei KvarQ, Mykrobe Predictor TB und TBProfiler musste der Quelltext editiert werden. Bei PhyResSE dagegen konnte die Mutationsliste (CSV-Datei), auf der die Resistenzvorhersage basiert, heruntergeladen werden und beliebig um weitere Mutationen ergänzt werden. Die erweiterte Mutationsliste konnte nach der Bearbeitung erneut hochgeladen werden und die aktuelle Analyse wurde automatisch an die neue Mutationsliste angepasst.

4.2 Evaluierung der phylogenetischen Klassifizierung

Traditionell wird die phylogenetische Klassifizierung von MTBK-Stämmen mit klassischen Typisierungsverfahren, wie Spoligotypisierung oder MIRU-VNTR Typisierung vorgenommen (Kapitel 1.5). Ein alternatives Verfahren ist die Analyse basierend auf phylogenetisch informativen SNPs, die zum Beispiel aus NGS-Daten extrahiert werden können [32, 82, 196]. Allerdings gibt es zurzeit noch kein allgemein gültiges Schema, sodass die getesteten Programme ihre phylogenetische Klassifizierung jeweils auf unterschiedliche SNPs-Datensätze stützen.

CASTB führt die traditionellen, genetischen Laborverfahren (Spoligotypisierung, MIRU-VNTR Typisierung, u. A.) virtuell anhand der NGS-Daten durch. KvarQ verwendet jeweils drei SNPs pro phylogenetischer Hauptlinie, die aus den Studien von Comas *et al.* 2009 und Stucki *et al.* 2012 stammen [34, 196]. Die Typisierung von Mykrobe Predictor TB basiert ebenfalls auf dem SNP-Schema von Stucki *et al.* 2012 [196]. Basierend auf diesem SNP-Schema, kann die Einteilung der Stämme bei diesen Programmen nur in die sieben Hauptlinien des MTBK erfolgen (Tabelle 4.3). Im Gegensatz dazu verwenden PhyResSE und TBProfiler zusätzlich spezielle Mutationen aus den Studien von Homolka *et al.* 2012 [82] und Coll *et al.* 2014 [32], die eine feinere Klassifikation in weitere MTBK-Untergruppen ermöglichen.

In Tabelle 4.3 ist die phylogenetische Klassifizierung der MTBK-Stämme des Sierra Leone Datensatzes mittels der klassischen MIRU-VNTR Typisierung sowie mittels der verschiedenen Programmen zur automatisierten Analyse von NGS-Daten dargestellt. Die Klassifizierung in die Hauptlinien des MTBK war für alle fünf getesteten Programme bis auf zwei Ausnahmen vergleichbar. Ein Stamm der Linie LAM wurde von CASTB fälschlicherweise als *M. bovis* (von Natur aus PZA-resistent) klassifiziert. Außerdem konnte Mykrobe Predictor TB sechs Stämme der Hauptlinie 5 (West African 1) nicht klassifizieren (Tabelle 4.3).

Mit allen Programmen war es möglich, elf zuvor nicht klassifizierte Stämme, der Euro-Amerikanischen Superlinie zuzuordnen. PhyResSE und TBProfiler konnten diese Stämme noch weiter den MTBK-Untergruppen 4.1.1.3, 4.1.2.1 (Haarlem), 4.6.1 und 4.8 zuordnen (Tabelle 4.3).

Insgesamt konnten KvarQ, PhyResSE, TBProfiler und für die phylogenetische

Klassifizierung der Sierra Leone Stämme 100 % Übereinstimmung mit den klassischen Typisierungsverfahren erreichen. CASTB zeigte in diesem Test eine Übereinstimmung von 98,9 % und Mykrobe Predictor TB hatte eine Übereinstimmung von 93,4 %.

4.3 Evaluierung der Resistenzvorhersage

Wie bei der phylogenetischen Klassifizierung steht für die Resistenzdetektion keine allgemein gültige Mutationsdatenbank zur Verfügung. Aus diesem Grund unterschieden sich die Resistenzvorhersage der getesteten Programme basierend auf den individuellen Mutationslisten. Die Mutationslisten, soweit diese von den Autoren zur Verfügung gestellt wurden, sind im Anhang D aufgeführt.

Die Resistenzdetektion von CASTB basiert laut Autoren auf Resistenzmarkern der Studien von Ando *et al.* 2010/2014 und Sekiguchi *et al.* 2007 [7,8,181]. Da die detektierten Mutationen von CASTB nicht ausgegeben werden, werden im Folgenden Aussagen über die detektierten Mutationen von CASTB mit Vorbehalt gegeben und basieren auf dem Vergleich mit der Mutationsdetektion der anderen Programme.

In der Studie von Feuerriegel *et al.* aus dem Jahr 2012 konnten in dem Datensatz aus Sierra Leone Resistenzmutationen für die Antibiotika INH, RMP, SM, EMB und PZA detektiert werden [55].

Der Vergleich in dieser Arbeit zeigte, dass selbst für die Resistenzprofile der wichtigsten Erstrangantibiotika INH und RMP erhebliche Diskrepanzen zwischen den Programmen beobachtet werden konnten. Als Goldstandardmethode zur Evaluierung der Resistenzvorhersage der verschiedenen Programme wurden die Sanger-Sequenzierungsergebnisse der Stämme aus Sierra Leone verwendet. Die Ergebnisse sind in Tabelle 4.4 dargestellt. In Tabelle 4.4 ist zu erkennen, dass nur wenige SNPs von allen Programmen gleichermaßen detektiert wurden (schwarze Zeilen). Zehn der resistenzvermittelnden Mutationen, die im Sierra Leone Datensatz mit der Sanger-Sequenzierung identifiziert worden sind, wurden von allen Programmen detektiert (Tabelle 4.4). Insgesamt war die Resistenzvorhersage für RMP vergleichbar für alle Programme. Es konnten lediglich zwei Ausnahmen beobachtet werden, die zur geringeren Sensitivität für CASTB ($\text{Sens}_{\text{RMP}} = 94\%$) und KvarQ ($\text{Sens}_{\text{RMP}} = 94\%$) führten (Tabelle 4.4 und Tabelle 4.5). In Tabelle 4.5 sind

die resultierenden Sensitivitäten und Spezifitäten der verschiedenen Programme für die einzelnen Antibiotika im Vergleich zur Sanger-Sequenzierung dargestellt. TBProfiler hat sieben Stämme, die mittels Sanger-Sequenzierung als INH-sensible vorhergesagt wurden, als resistent klassifiziert. Alle Stämme wurden von TBProfiler anhand des SNPs *kasA* G269S als INH-resistent interpretiert, obwohl dieser SNP als spezifische Variante für die Linie LAM/4.3.3 bekannt ist, die keine Resistenz vermittelt [198]. Aus diesem Grund war die INH-Spezifität für TBProfiler mit $\text{Spez}_{\text{INH}} = 85\%$ im Vergleich gering (Tabelle 4.5).

Besonders für die Antibiotika EMB, SM und PZA waren Sensitivität und Spezifität der Resistenzvorhersage für die Programme sehr unterschiedlich. Für EMB beispielsweise haben CASTB und KvarQ lediglich acht der insgesamt 16 in der Testkollektion vorhandenen resistenzvermittelnden SNPs im *embABC*-Operon korrekt detektiert (Tabelle 4.4). Mykrobe Predictor TB detektierte sogar nur sieben Stämme korrekt als EMB-resistent. Im Gegensatz dazu stimmte die EMB-Resistenzvorhersage von PhyResSE und TBProfiler sehr gut mit den Sanger-Sequenzierungsergebnissen überein (Tabelle 4.5), obwohl jeweils ein resistenzvermittelnder SNP (*embB* D354A aus [25] bzw. *embB* Trp332Arg aus [55]) nicht in der jeweiligen Mutationsliste enthalten war (Tabelle 4.4).

Für SM wurden ebenfalls nur geringe Vorhersagewerte mit Sensitivitäten zwischen 30% für CASTB und 86% für PhyResSE berechnet. CASTB konnte in diesem Test ausschließlich Stämmen korrekt als SM-resistent vorhersagen, bei denen der SNP *rpsL* K43R detektiert wurde. Dadurch hat CASTB elf der laut Sanger-Sequenzierung 22 SM-resistenten Stämme korrekt vorhergesagt. Die anderen Programme verwendeten zusätzliche SNPs in *rrs*, *rpsL* und *gidB* (nur PhyResSE) für die SM-Resistenzvorhersage (Tabelle 4.4). Auf diese Weise erzielte PhyResSE die höchste Sensitivität ($\text{Sens}_{\text{SM}} = 86\%$). Dennoch konnte PhyResSE sechs der 17 mit Sanger-Sequenzierung detektierten *gidB*-Mutationen nicht detektieren. Zum einen hatte PhyResSE den SNP *gidB* G69D ($n = 2$) nicht in der Liste der resistenzvermittelnden Mutationen [208] und zum anderen konnten Indels ($n = 4$) nicht korrekt identifiziert werden. Diese wurden zwar ausgegeben, aber in der verwendeten Version 1.0 nicht für die Resistenzvorhersage verwendet.

Die Werte für die Resistenzvorhersage von PZA waren bei allen getesteten Programmen besonders gering (Tabelle 4.5). CASTB und KvarQ detektierten jeweils nur fünf der insgesamt zwölf PZA-resistenten Stämme korrekt im Vergleich zu den Sanger-Sequenzierungsergebnisse (Tabelle 4.4). PhyResSE hatte in die-

sem Vergleich die meisten Übereinstimmungen mit den Ergebnissen der Sanger-Sequenzierung des Gens *pncA* ($n = 9$) und damit die höchste Sensitivität in diesem Test ($\text{Sens}_{\text{PZA}} = 75\%$; Tabelle 4.5). TBProfiler hat sieben SNPs im Vergleich zu den Ergebnissen der Sanger-Sequenzierung korrekt detektiert (Tabelle 4.4). Der Vollständigkeit halber wurde für alle Programme die jeweilige Sensitivität und Spezifität für die einzelnen Antibiotika basierend auf der phänotypischen Resistenztestung berechnet (Anhang E, Tabelle E.1).

4 Evaluierung automatischer NGS-Analyseprogramme mit Hilfe eines Validierungsdatensatzes

Tabelle 4.3: Phylogenetische Klassifizierung der MTBK-Stämme des Sierra Leone Datensatzes mit den verschiedenen Programmen zur automatisierten Analyse von NGS-Daten. Modifiziert aus: [179]

Linie ¹	#	CASTB	KvarQ	Mykrobe Predictor TB	PhyResSE	TBProfiler
EAI	4	Linie 1 (Indo-Oceanic Linie)	Linie 1	EAI	EAI/ Linie 1.1.1	Linie 1.1.1
Beijing	4	Linie 2 (East-Asian Linie)	Linie 2/ Beijing	Beijing/ East Asia	Beijing/ Linie 2.2.1	Linie 2.2.1
Ghana	1	Linie 4 (Euro-Am)	Linie 4	Euro-Am	Ghana/ Linie 4.1	Linie 4.1
Sierra Leone-1	7	Linie 4 (Euro-Am)	Linie 4	Euro-Am	Euro-Am/ Linie 4.1	Linie 4.1
X-type	1	Linie 4 (Euro-Am)	Linie 4	Euro-Am	Euro-Am/ Linie 4.1.1.1	Linie 4.1.1.1
	1	Linie 4 (Euro-Am)	Linie 4	Euro-Am	Euro-Am/ Linie 4.1.1.3	Linie 4.1.1.3
Haarlem	11	Linie 4 (Euro-Am)	Linie 4	Euro-Am	Haarlem/ Linie 4.1.2.1	Linie 4.1.2.1
	9	Linie 4 (Euro-Am)	Linie 4	Euro-Am	LAM/ Linie 4.3.3	Linie 4.3.3
LAM	1	Linie <i>M. bovis</i> ²	Linie 4	Euro-Am	LAM/ Linie 4.3.3	Linie 4.3.3
	2	Linie 4 (Euro-Am)	Linie 4	Euro-Am	LAM/ Linie 4.3.4.1	Linie 4.3.4.1
	3	Linie 4 (Euro-Am)	Linie 4	Euro-Am	LAM/ Linie 4.3.4.2	Linie 4.3.4.2
S-type	3	Linie 4 (Euro-Am)	Linie 4	Euro-Am	S-type/ Linie 4.4.1.1	Linie 4.4.1.1
Cameroon	4	Linie 4 (Euro-Am)	Linie 4	Euro-Am	Cameroon/ Linie 4.6.2.2	Linie 4.6.2.2
Sierra Leone-2	9	Linie 4 (Euro-Am)	Linie 4	Euro-Am	Euro-Am/ Linie 4.8	Linie 4.8
WA 1	6	Linie 5 (WA 1)	Linie 5	Unbekannte Spezies ³	WA 1/ Linie 5	Linie 5
WA 2	14	Linie 6 (WA 2)	Linie 6	<i>M. africanum</i>	WA2/ Linie6	WA 2/ Linie 6
	3	Linie 4 (Euro-Am)	Linie 4	Euro-Am	Euro-Am/ Linie 4.1	Linie 4.1
	3	Linie 4 (Euro-Am)	Linie 4	Euro-Am	Euro-Am/ Linie 4.1.1.3	Linie 4.1.1.3
Nicht klassifiziert	1	Linie 4 (Euro-Am)	Linie 4	Euro-Am	Haarlem/ Linie 4.1.2.1	Linie 4.1.2.1
	1	Linie 4 (Euro-Am)	Linie 4	Euro-Am	Euro-Am/ Linie 4.6	Linie 4.6.1
	3	Linie 4 (Euro-Am)	Linie 4	Euro-Am	Euro-Am/ Linie 4.8	Linie 4.8

Abkürzungen: EAI:East African/Indian Ocean; Euro-Am: Euro-Amerikanische Superlinie; LAM: Latin-American Mediterranean; WA: West African

¹ Klassifizierung mittels MIRC-VNTR Typisierung

² CASTB klassifiziert einen Stamm der Linie LAM (Linie 4.3.3) als *M. bovis*

³ Mykrobe Predictor TB kann Bakterien der West African 1 Linie nicht klassifizieren

Tabelle 4.4: Detektierte Mutationen der evaluierten Programme zur automatisierten Analyse von NGS-Daten bei der Untersuchung des Sierra Leone Datensatzes. Unter # ist die Anzahl der jeweiligen Mutation im Sierra Leone Datensatz aufgeführt. Zellfarbe weiß: Mutation nicht in der Mutationsliste, auf der die Resistenzvorhersage basiert; Schwarz: Mutation in der Mutationsliste, auf der die Resistenzvorhersage basiert. Modifiziert aus: [179]

Gen	Mutation	#	CASTB	KvarQ	Mykr.Pr.TB	PhyResSE	TBProfiler
<i>ahpC</i>	C-57T	1					
<i>embA</i>	C-12T	1					
<i>embB</i>	M306V	1					
	M306I	6					
	W332R	1					1
	D354A	1					
	G406A	1			1		
	G406D	1		1	2		
	Q497K	2		1			
	Q497P	1		1		1	
	H1002R	2					
<i>gidB</i>	G34A	1					
	V65G	1					
	G69D	2					
	V88A	2					
	L91P	1					
	Q127-	1					
	A138V	2					
	A200E	3					
	Indels	4					
<i>inhA</i>	C-15T	3					
	G-17T	1					
	S94A	1					
<i>kasA</i>	G269S	10					
<i>katG</i>	W300C	1					1
	S302R	1					
	S315T	22					
<i>pncA</i>	T-11C	1					
	I5T	1					1
	T47A	2					
	P62S	1				1	1
	T87M	1					
	A146E	1					1
	G162D	1					
	L172P	1		1			
	V180F	1		1			
	10 bp-Insertion (143)	1					
	1 bp-Insertion (64)	1					
<i>rpoB</i>	T400A	1					
	L430P	1					
	D435Y	2					
	H445R	3					
	H445Y	4					
	S450L	8					
	L452P	2					
<i>rpsL</i>	K43R	11					
	K88R	9					
<i>rrs</i>	C513T	2		3			

Abkürzungen: Mykr.Pr.TB-Mykrobe Predictor TB.

¹ Position in der Mutationsliste, aber mit anderem Basenaustausch.

² Aminosäureaustausch in Mutationsliste, aber nicht detektiert.

³ Ausgabe als S171S (interpretiert den Aminosäureaustausch).

Tabelle 4.5: Sensitivität und Spezifität der Resistenzvorhersage verglichen mit der Sanger-Sequenzierung für die einzelnen Antibiotika im Test. Berechnet mit R Version 3.1.2. Modifiziert aus: [179]

Antib. #R	Sanger #S	CASTB		KvarQ		Mikrobe Sens	Predictor TB		PhyResSE		TBProfiler	
		Sens	Spez	Sens	Spez		Sens	Spez	Sens	Spez	Sens	Spez
INH	28	63	0,83 (0,64; 0,94)	1,00 (0,91; 1,00)	0,83 (0,64; 0,94)	1,00 (0,91; 1,00)	0,86 ² (0,68; 0,96)	1,00 (0,91; 1,00)	1,00 (0,83; 1,00)	1,00 (0,91; 1,00)	0,93 (0,77; 0,99)	0,85 ¹ (0,74; 0,93)
RMP	18	73	0,94 (0,73; 1,00)	1,00 (0,93; 1,00)	0,94 (0,73; 1,00)	1,00 (0,93; 1,00)	1,00 (0,74; 1,00)	0,99 ^{2,3} (0,93; 1,00)	1,00 (0,74; 1,00)	0,99 (0,93; 1,00)	1,00 (0,74; 1,00)	0,99 (0,93; 1,00)
SM	37	54	0,30 (0,16; 0,47)	1,00 (0,90; 1,00)	0,57 (0,39; 0,73)	1,00 (0,90; 1,00)	0,57 (0,39; 0,73)	1,00 (0,90; 1,00)	0,86 (0,71; 0,95)	1,00 (0,90; 1,00)	0,57 (0,39; 0,73)	1,00 (0,90; 1,00)
EMB	15	76	0,53 (0,27; 0,79)	1,00 (0,93; 1,00)	0,53 (0,27; 0,79)	1,00 (0,93; 1,00)	0,47 ³ (0,21; 0,73)	0,99 (0,93; 1,00)	0,93 (0,68; 1,00)	1,00 (0,93; 1,00)	0,93 (0,68; 1,00)	0,98 (0,93; 1,00)
PZA	11	80	0,42 (0,15; 0,72)	1,00 (0,93; 1,00)	0,42 (0,15; 0,72)	1,00 (0,93; 1,00)	n.v. n.v.	n.v. n.v.	0,75 (0,43; 0,95)	1,00 (0,93; 1,00)	0,58 (0,28; 0,85)	1,00 (0,93; 1,00)

Abkürzungen: #-Anzahl, Antib.-Antibiotikum, EMB-Ethambutol, INH-Isoniazid, n.v.-nicht verfügbar, PZA-Pyrazinamid, R-resistent, RMP-Rifampicin, S-sensibel, Sens-Sensitivität (richtig positive Rate), SM-Streptomycin, Spez-Spezifität (richtig negativ Rate).

¹ *Kas4* wurde mittels Sanger-Sequenzierung nicht sequenziert

² Nur interpretierte Positionen eingeschlossen (keine niederfrequenten Mutationen < 15% alternativen Allel)

³ Niederfrequente SNPs, die initial nicht mit Sanger detektiert wurden, werden als Wildtyp behandelt

4.4 Detektion von Mischinfektionen

Im Folgenden soll die Detektion von Mischinfektionen überprüft werden. Dafür wurden die NGS-Daten der artifiziellen Mischungen Mix50, Mix70 und Mix90 (Tabelle 2.3) mit den jeweiligen Programmen analysiert.

Die Ergebnisse unterschieden sich stark für die jeweiligen Programme (Tabelle 4.6). CASTB ist theoretisch auf Grund der separaten Ausgabe verschiedener Typisierungsmethoden, wie zum Beispiel der Beijing-Subtypisierung und der LSP-Typisierung, in der Lage, Mischungen zu erkennen, in denen ein Stamm der Linie Beijing mit einem Stamm einer anderen phylogenetischen Linie des MTBK gemischt ist. In diesem Test waren die Ergebnisse von CASTB jedoch nicht konsistent. Zweimal wurde der Stamm der Linie Beijing der modernen und einmal der ancestralen Familie zugeordnet (Tabelle 4.6). Darüber hinaus wurde die sensible Mischung Mix90 fälschlicherweise als PZA-resistent klassifiziert. KvarQ klassifi-

Tabelle 4.6: Phylogenetische Klassifizierung der artifiziellen Mischinfektionen Mix50, Mix70 und Mix90 mit den getesteten Programmen.

Programm	Mix50	Mix70	Mix90
CASTB	Linie 2 (ancestral) Linie 4	Linie 2 (modern)	Linie 2 (modern) Linie 6 PZA-Resistenz
KvarQ	Linie 2 (50 %) Linie 4 (50 %)	Linie 2 (70 %) Linie 4 (30 %)	Linie 2 (90 %) Linie 4 (10 %)
Mykr. Pred. TB	Linie Beijing/Ost Asien	Linie Beijing/Ost Asien	Linie Beijing/Ost Asien
PhyResSE	Linie 2.2.1 (51 %) Linie 4.1.2.1 (43 %)	Linie 2.2.1 (73 %) Linie 4.1.2.1 (25 %)	Linie 2.2.1 (90 %)
TBProfiler	Linie 2.2.1 Linie 4.1.2.1	Linie 2.2.1 Linie 4.1.2.1	Linie 2 Linie 4.1.2

Abkürzungen: Mykr. Pred. TB- Mykrobe Predictor TB, PZA-Pyrazinamid.

zierte alle artifiziellen Mischproben Mix50, Mix70 und Mix90 initial als Stämme der Linie Beijing, gab allerdings jeweils einen Hinweis auf uneinheitliche Ergebnisse für Positionen, die spezifisch für die Linie 4 sind. Mit Hilfe zusätzlicher Allelfrequenzinformationen konnten Rückschlüsse auf das Mischverhältnis gezogen werden (Tabelle 4.6).

Mykrobe Predictor TB hat alle drei Mischungen Mix50, Mix70 und Mix90 als Stämme der Linie Beijing klassifiziert und somit keine der Mischungen korrekt identifiziert.

Die phylogenetische Analyse von PhyResSE besteht aus einer initialen Klassifizierung basierend auf einem Entscheidungsbaum nach Homolka *et al.* 2012 [82] und der anschließenden erklärenden Ergebnistabelle mit den linienspezifischen SNPs. Der Entscheidungsbaum ordnete alle drei artifiziellen Mischungen Mix50, Mix70 und Mix90 (Tabelle 2.3) der Linie Beijing zu. Erst die Auswertung der detaillierten Ergebnistabelle ließ Rückschlüsse auf die jeweiligen Mischungen zu (Tabelle 4.6).

TBProfiler detektierte für alle drei Mischungen Mix50, Mix70 und Mix90 spezifische SNPs beider Linien (Linie 2 und Linie 4), allerdings konnten keine Aussagen über das Mischungsverhältnis der beiden Stämme abgeleitet werden (Tabelle 4.6). Zusammenfassend bietet KvarQ im Vergleich die beste Lösung zur Erkennung von Mischinfektionen zweier verschiedener phylogenetischer Linien in diesem Vergleich. Eine Mischung zweier Untergruppen der gleichen Hauptlinie war allerdings auch mit KvarQ nicht möglich (Daten nicht gezeigt), sodass keines der Programme eine optimale Lösung zur Detektion von Mischinfektionen lieferte. Es wird eine neue Methode benötigt, die sowohl zwischen Mischungen unterschiedlicher phylogenetischer Linien, als auch zwischen Mischungen derselben phylogenetischen Linie zuverlässig unterscheiden kann und eine Schätzung des Verhältnisses der unterschiedlichen Stämme erlaubt.

4.5 Detektion von Heteroresistenzen

In dem Sierra Leone Datensatz konnte die NGS-Probe 420-04 (ENA-Zugriffsnummer: ERS456829) als heteroresistent klassifiziert werden (Abbildung 4.1), obwohl diese Heteroresistenz während der Standardprozedur der Sanger-Sequenzierung nicht detektiert wurde. Im folgenden Abschnitt werden die Resistenzvorhersagen der Programme für die NGS-Daten dieses Stammes und der artifiziellen, heteroresistenten Proben (Kapitel 2.4) dargestellt.

Die Qualität der Resistenzvorhersage der Programme hängt von der Frequenz des alternativen Allels ab.

CASTB konnte zum Beispiel den niederfrequenten SNP der NGS-Probe 420-04 (ENA-Zugriffsnummer: ERS456829; Abbildung 4.1) mit circa 20 % alternativem Allel nicht detektieren. KvarQ zeigte für die Probe 420-04 (ENA-Zugriffsnummer: ERS456829; Abbildung 4.1) immerhin an, dass Positionen mit nicht eindeutigem Ergebnis („mixed coverage“) vorhanden sind. Eine genaue Analyse der von KvarQ zur Verfügung gestellten Abdeckungsgrafik der Position 761161 zeigte, dass zwölf der insgesamt 71 ausgerichteten NGS-Sequenzfragmente (17 %) die alternative Base „C“ zeigten (*rpoB* L452P).

Mykrobe Predictor TB ist ein Programm, das speziell für die Detektion von niederfrequenten Varianten implementiert wurde [33]. MykrobePredictor TB, PhyResSE und TBProfiler klassifizierte die NGS-Probe 420-04 (ENA-Zugriffsnummer: ERS456829; Abbildung 4.1) korrekt als RMP-resistent basierend auf der Mutation *rpoB* H452P. Mit Mykrobe Predictor TB und PhyResSE (Abbildung 4.1) konnte eine Frequenz von etwa 20 % abgeleitet werden, während von TBProfiler keine Frequenzinformation gegeben wurde.

Das heißt niederfrequente SNPs mit einer Frequenz von über 20 % alternativem Allel können mit den getesteten Programmen (außer CASTB) zuverlässig detektiert werden.

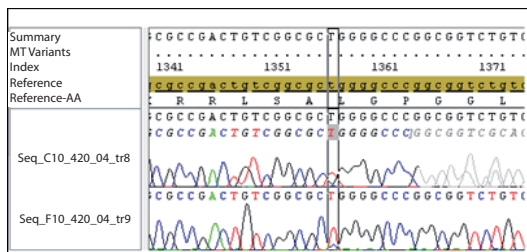
Die Ergebnisse der verschiedenen Programme für die artifiziellen, heteroresistenten NGS-Proben sind in Tabelle 4.7 dargestellt. Für CASTB konnten keine Ergebnisse erstellt werden, da die FastQ-Dateien zu groß zum Hochladen auf der Webseite waren. Für die Programme KvarQ, Mykrobe Predictor TB und TBProfiler wurden jeweils die Kommandozeilenanwendungen zur Analyse verwendet.

Von den artifiziellen Mischungen, die den SR4k-Selektionsklon enthielten (Tabelle 2.3), konnte nur Mykrobe Predictor TB den gesuchten SNP bei den Pro-

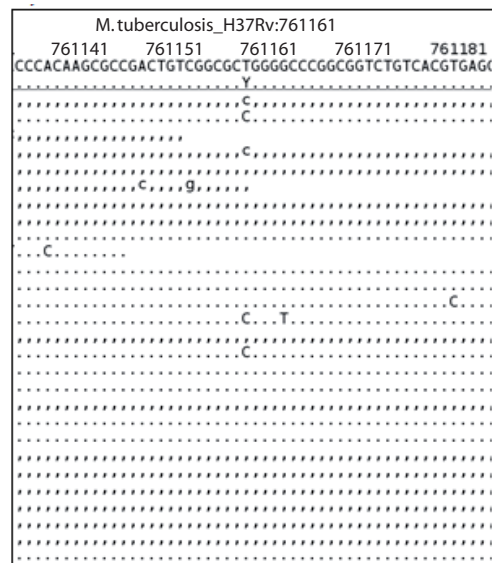
4 Evaluierung automatischer NGS-Analyseprogramme mit Hilfe eines Validierungsdatensatzes

Position <small>hover:nt_counts link:map'd_reads</small>	Reference <small>in genome direction</small>	Sample <small>in genome direction (%reads)</small>	Quality <small>[+]</small>	Gene Name <small>link:TubercuList</small>	Dir. <small>[x]</small>	AA change <small>[x]</small>	Codon <small>change</small>	Variant <small>position gene start</small>	Antibiotic <small>[x]</small>	High <small>Confidence SNP</small>
761161	T	C _(19%)	204,77	rpoB	+	Leu452Pro	ctg/ccg	1355	rifampicin (RMP)	yes

(a) PhyResSE Ausgabe



(b) Sanger Ausgabe



(c) SAMtools Visualisierung

Abbildung 4.1: Position 761161 der heteroresistenten NGS-Probe 420-04 (ENA-Zugriffsnummer: ERS456829) des Sierra Leone Datensatz mit Hilfe von PhyResSE und der Sanger-Sequenzierung [56]. **a** Screenshot der Resistenzvorhersage von PhyResSE für den Stamm 420-04 (ENA-Zugriffsnummer: ERS456829). *RpoB* SNP mit 19% alternativem Allel an Position 761161. **b** Sanger-Chromatogramm für den Stamm 420-04 (ENA-Zugriffsnummer: ERS456829) an der Position 761161. Ein schwaches Signal, das während der Standardprozedur nicht detektiert wurde. **c** Darstellung der Ausrichtung der NGS-Sequenzfragmente des Stammes 420-04 (ENA-Zugriffsnummer: ERS456829) zur Referenzsequenz an der Position 761161; etwa 20% der NGS-Sequenzfragmente zeigten die alternative Base.

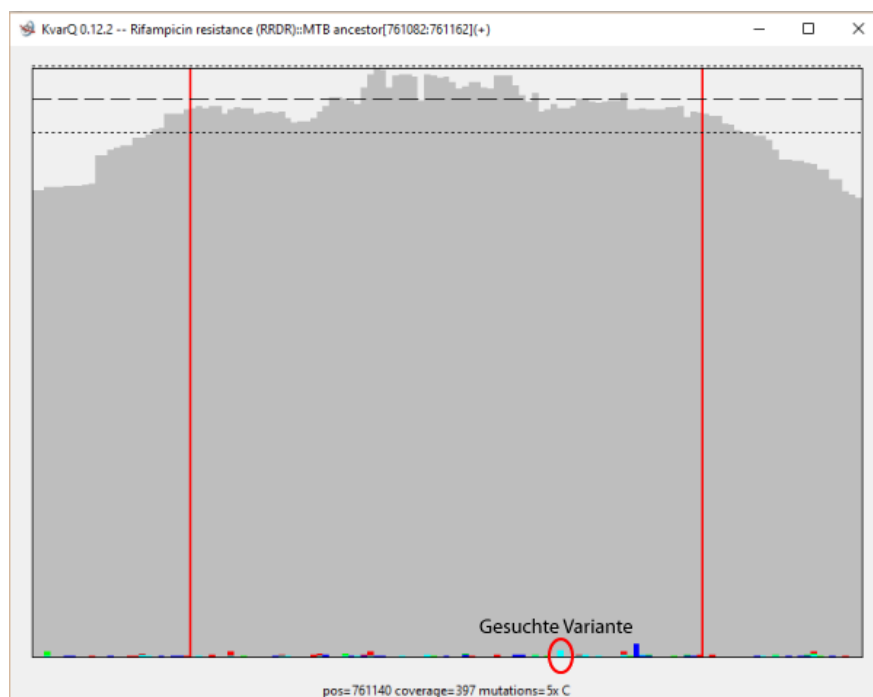


Abbildung 4.2: Abdeckungsgrafik von KvarQ für die Kernregion RRDR des *rpoB*-Gens für die NGS-Probe SR4k_rpoB526_1. Es konnte eine Mutation an der Position 761140 (*rpoB* 526) beobachtet werden.

ben SR4k_rpoB526_5 und SR4k_rpoB526_10 detektieren (Tabelle 4.7), allerdings wurde auf Grund der geringen Frequenz keine Resistenz interpretiert. KvarQ, PhyResSE und TBProfiler haben keine Probe korrekt als RMP-resistent vorhergesagt. Es gab weder eine Angabe der Mutation, noch einen Hinweis auf eine gemischte Allelpopulation (KvarQ) an den betreffenden Stellen. Eine Analyse der Abdeckungsgrafiken (KvarQ), beziehungsweise der Visualisierung der zur Referenzsequenz ausgerichteten NGS-Sequenzfragmente (PhyResSE), zeigte, dass die entsprechenden Mutationen zwar vorhanden waren, aber ein Grenzwert zur SNP-Detektion nicht überschritten wurde. Für die NGS-Probe SR4k_rpoB526_10 wurden an der Position 761140 zum Beispiel lediglich etwa 8% alternatives Allel gezählt. Außerdem bestand in den Abdeckungsgrafiken von KvarQ keine Möglichkeit zum Beispiel anhand von Qualitätswerten die gesuchten Mutationen von Fehlern zu unterscheiden (Abbildung 4.2).

Für die NGS-Proben, die den Selektionsklon mit einer Mutation bei *rpoB* 531 enthielten (Tabelle 2.3), waren die Ergebnisse ähnlich, allerdings hat KvarQ bei der NGS-Probe SR1a_rpoB531_10 auf eine gemischte Allelpopulation hingewiesen

Tabelle 4.7: Ergebnisse der automatisierten NGS-Analyseprogramme für die artifiziellen, heteroresistenten NGS-Proben.

Probe	KvarQ	Mykr. Pred. TB	PhyResSE	TBProfiler
SR4k_rpoB526_1	-	-	-	-
SR4k_rpoB526_5	-	5,5 % <i>rpoB</i> H445X	-	-
SR4k_rpoB526_10	-	8,5 % <i>rpoB</i> H445X	-	-
SR1a_rpoB531_1	-	-	-	-
SR1a_rpoB531_5	-	5,3 % <i>rpoB</i> S450X	-	-
SR1a_rpoB531_10	10 % T 761155	10,3 % <i>rpoB</i> S450X	-	-

Abkürzungen: Mykr. Pred. TB-Mykrobe Predictor TB

(Tabelle 4.7). Eine genauere Untersuchung der Kernregion RRDR der NGS-Probe SR1a_rpoB531_10 zeigte für die Position 761155 (*rpoB* 531) 10 % alternatives Allel.

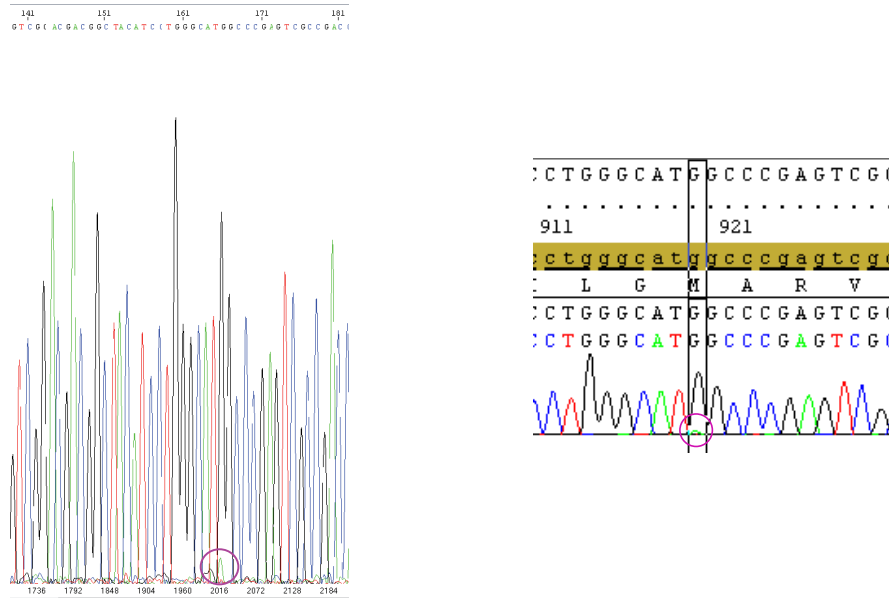
Zusätzlich wurden mit Mykrobe Predictor TB im Sierra Leone NGS-Datensatz sieben Stämme mit einer niederfrequenten RMP-Resistenzmutationen (< 4 % alternatives Allel) und ein Stamm (8867-03; ENA-Zugriffsnummer: ERS457115) mit einer niederfrequenten EMB-Resistenzmutation (13 % alternatives Allel) detektiert. Für diese Stämme war das phänotypische DST-Ergebnis sensitiv gegenüber RMP, beziehungsweise EMB und die initiale Untersuchung der Sanger-Sequenzen ergab keine resistenzvermittelnden Mutationen.

Die sieben RMP-Resistenzmutationen wurden nicht in der Hauptausgabe von Mykrobe Predictor TB (erstes Fenster) angegeben, sondern im Evidenz-Teil (letztes Fenster). Zur Bestimmung der Qualität dieser Varianten wurden die entsprechenden FastQ-Dateien mit dem Algorithmus „Stampy“ zur Referenzsequenz *M. tuberculosis* H37Rv ausgerichtet und mit einem Programm zur visuellen Durchsicht der Referenzgenom-Ausrichtung, dem Integrative Genome Viewer (IGV), untersucht. Alle alternativen Basen, unabhängig von der NGS-Probe und der Position, hatten einen Qualitätswert unter 15, das heißt die Wahrscheinlichkeit, dass diese Basen fehlerhaft von dem Sequenziergerät bestimmt wurden, lag bei über 3 % (Tabelle 2.1).

Der Stamm 8867-03 (ENA-Zugriffsnummer: ERS457115) wurde von Mykrobe Predictor TB als EMB-resistent klassifiziert (erstes Fenster). Die Sanger-Sequenz des *embB*-Gens des Stammes 8867-03 (ENA-Zugriffsnummer: ERS457115) wurde erneut untersucht; es konnte sowohl auf dem Chromatogramm, als auch auf dem Elektropherogramm ein schwaches Signal für die resistenzvermittelnde Variante

beobachtet werden (Abbildung 4.3).

Der Stamm 8867-03 (ENA-Zugriffsnummer: ERS457115) wurde von PhyRes-



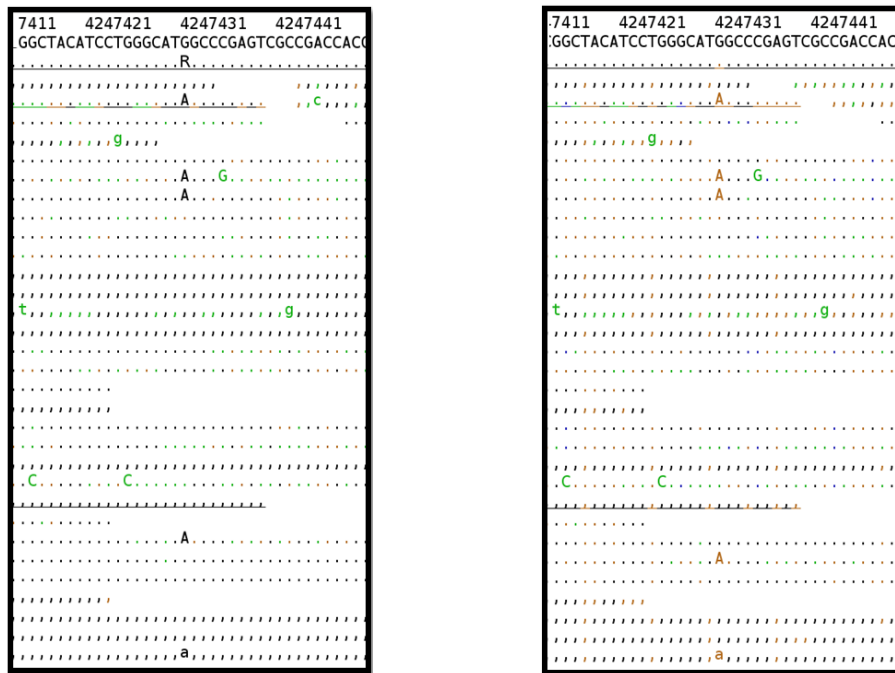
(a) Sanger-Electropherogramm
der Genomposition 4247431

(b) Sanger-Chromatogramm
der Genomposition 4247431

Abbildung 4.3: Sanger-Ergebnis der niederfrequenten EMB-Resistenzmutation an Genomposition 4247431 der NGS-Probe 8867-03 (ENA-Zugriffsnummer: ERS457115). Sowohl auf dem Chromatogramm (b), als auch auf dem Electropherogramm (a) konnte anhand eines schwachen Signals das Allel „A“ statt des Referenzallels „G“ nachgewiesen werden.

SE zwar nicht als EMB-resistent klassifiziert. Eine detaillierte Untersuchung der NGS-Sequenz des Stammes 8867-03 (ENA-Zugriffsnummer: ERS457115) zeigte allerdings, dass die alternativen Basen in den NGS-Daten vorhanden waren und die NGS-Sequenzfragmente von BWA auch korrekt zur Referenzsequenz ausgerichtet wurden, aber die Basenqualität durch die Rekalibrierung der Basenqualitätswerte so herab gesetzt wurden, dass PhyResSE keine Variante detektierte (Abbildung 4.4).

Für die Detektion von niederfrequenten Resistenzmutationen war keines der getesteten Programme geeignet. Es wird ein neues Verfahren benötigt, das niederfrequente Varianten detektiert, bewertet und die Frequenz vorhersagt.



(a) Ohne Rekalibrierung der Basenqualitäts- (b) Mit Rekalibrierung der Basenqualitäts-
 werte werte

Abbildung 4.4: Ausrichtung der NGS-Sequenzfragmente an der Position 4247431 mit unterschiedlichen Analyseabfolgen. Visualisierung der Ausrichtung der NGS-Sequenzfragmente zur Referenzsequenz mit SAMtools tview [56]. Farbkodierung der Basenqualitätswerte: Blau 0-9; Grün 10-19; Gelb 20-29 und Schwarz ≥ 30 . Ein Punkt/Komma zeigt eine Übereinstimmung der Base auf einem vorwärts-/rückwärts-sequenziertem Fragment an. Buchstaben zeigen Nicht-Übereinstimmungen mit der Referenzbase an. **a** Ausrichtung der NGS-Sequenzfragmente zur Referenzsequenz ohne Rekalibrierung der Basenqualitätswerte. Es wurde ein SNP an der Position 4247431 G→A detektiert und folglich als EMB-resistant klassifiziert. **b** Ausrichtung der NGS-Sequenzfragmente zur Referenzsequenz nach der Rekalibrierung der Basenqualitätswerte. Es wurde kein SNP detektiert.

5 Heterogene Populationen

Heterogene Populationen können für den molekularbiologischen Nachweis von Resistenzen problematisch sein, wenn zum Beispiel für bestimmte Resistenzgene wildtyp und mutiertes Allel vorliegen, aber nur die wildtyp Variante nachgewiesen wird. Grund hierfür kann das Vorliegen verschiedene Populationen des gleichen Stamms vorliegen, die zum Beispiel durch diversifizierende Evolution unter Antibiotikatherapie entstanden sind, oder eine Infektion mit zwei verschiedenen Stämmen zum Beispiel durch Reinfektion mit einem weiteren MTBK-Stamm bei bestehender Infektion sein. Es konnte bereits gezeigt werden, dass keines der getesteten automatisierten NGS-Analyseprogramme heterogene Populationen optimal detektieren und den Anteil einzelner Populationen an der Gesamtpopulation akkurat bestimmen kann. Es werden daher neue Methoden zur Detektion von heterogenen Populationen benötigt. Im Rahmen der vorliegenden Arbeit wurden zwei Skripte zur Detektion von Mischinfektionen und niederfrequenten SNPs implementiert. In den folgenden Abschnitten wird die Entwicklung des „Mischungsdetektors“ basierend auf einem Clusteralgorithmus dargestellt und die Performance unter verschiedenen Bedingungen überprüft. Außerdem wird das Verfahren zur Detektion und Bewertung niederfrequenter SNPs entwickelt und auf klinischen Daten getestet.

5.1 Detektor für Mischinfektionen

Die Implementierung des Mischungsdetektor basiert auf der Annahme, dass sich in einer Mischung aus zwei unterschiedlichen MTBK-Stämmen eine größere Anzahl SNPs befinden, bei denen der Anteil des alternativen Allels an allen NGS-Sequenzfragmenten übereinstimmt und von 100 % abweicht. Das heißt bei einer

heterogenen Population mit zwei MTBK-Stämmen im Verhältnis 8:2 werden beispielsweise mehrere SNPs mit 20 % alternativem Allel (Stamm 1) sowie mehrere SNPs mit 80 % alternativem Allel (Stamm 2) und einige mit 100 % alternativem Allel (gemeinsame SNPs) erwartet. Demnach müssten sich in einem Grafen, der die Allelfrequenzen der detektierten SNPs darstellt, drei Häufungspunkte detektieren lassen.

Zur Überprüfung dieser Annahmen, sind in Abbildung 5.1 die Streudiagramme der Allelanteile der detektierten SNPs für die NGS-Proben Mix50, Mix70 und Mix90 (Tabelle 2.3) sowie für eine Kontrolle dargestellt. Mit roten Kreisen wurden Bereiche markiert, in denen visuell eine Häufung von Datenpunkten beobachtet werden konnte. Es ist zu erkennen, dass für die NGS-Proben Mix70 und die Kontrolle anhand der Häufungspunkte des Grafen die korrekten Frequenzen der Populationen abgeleitet werden können und somit die Grundannahme zur Erstellung des Mischungsdetektors optimal abgebildet wird. Ein etwas anderes Bild ergab sich für die Proben Mix50 (Abbildung 5.1a) und Mix90 (Abbildung 5.1c). Für die Probe Mix50 konnten zwar drei Häufungspunkte beobachtet werden, allerdings überschneiden sich bei 50 % alternativem Allel die SNPs beider enthaltenen Populationen. Wohingegen der Häufungspunkt unterhalb von 20 % alternativem Allel in Abbildung 5.1a auf Grund des bekannten Mischungsverhältnisses der Probe Mix50 als Rauschen identifiziert werden konnte. Bei der Untersuchung der Probe Mix90 konnten nur zwei Häufungspunkte beobachtet werden. Ein kleiner bei ungefähr 10 % alternativem Allel und ein größerer mit einem Schwerpunkt bei ungefähr 90 % alternativem Allel. Die SNPs, die in beiden Populationen auftraten, ließen sich hier optisch nicht von den SNPs der 90 % Population differenzieren.

Dennoch entsprechen die Abbildungen größtenteils der Grundannahme zur Erstellung des Mischungsdetektors, sodass im Folgenden eine automatisierte Erkennung der Häufungspunkte als Ansatz für die Erstellung eines Mischungsdetektors verwendet wurde.

Die Anforderungen an den Algorithmus lassen sich auf folgende Punkte bringen:

- Einteilung der Datenpunkte in Gruppen
- Automatisierte Detektion von Häufungspunkten
- Robust gegenüber Rauschen/Ausreißern

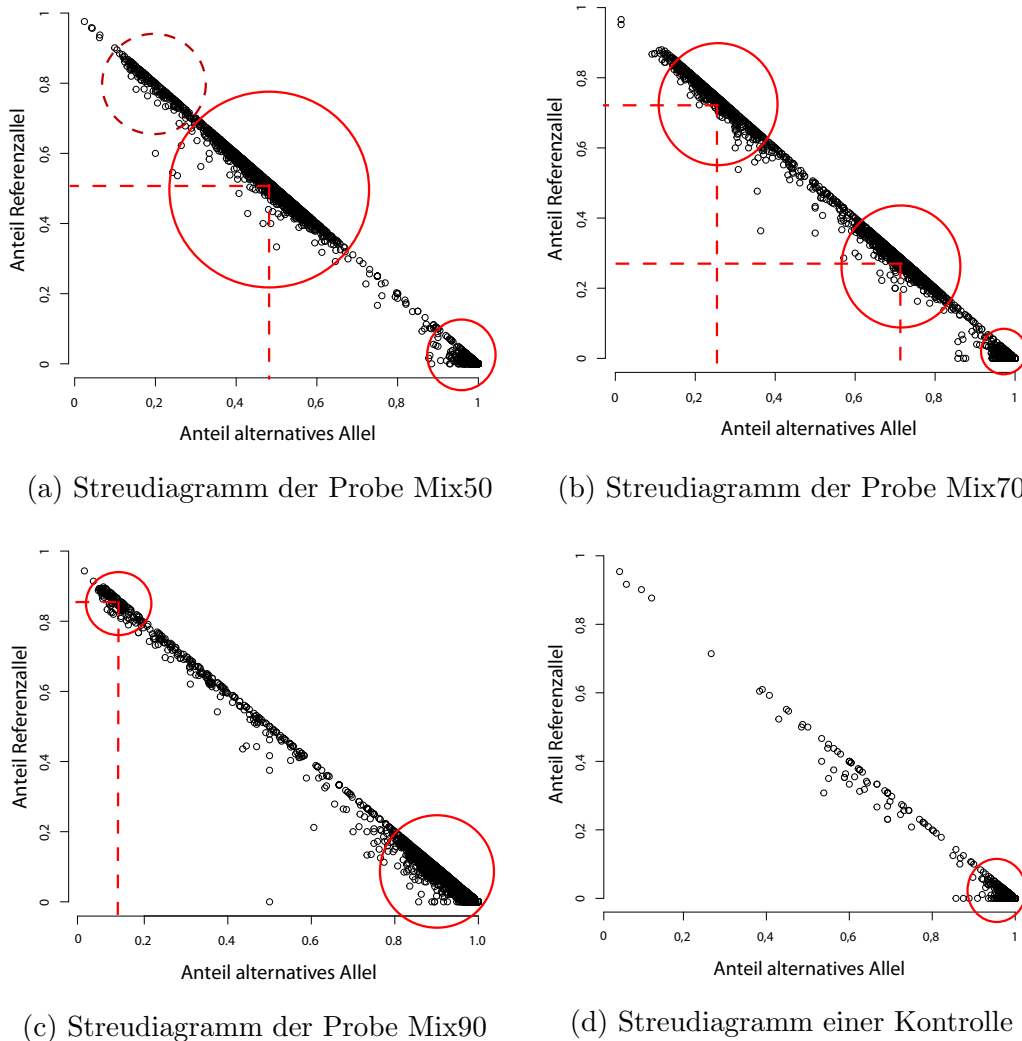


Abbildung 5.1: Streudiagramme der Allelfrequenzen der detektierten SNPs für die NGS-Proben Mix50 (a), Mix70 (b), Mix90 (c) und eine Kontrolle (d). Mit roten Kreisen sind Anhäufungen von Datenpunkten markiert. Darstellung mit R Version 3.0.2. **a** Es konnten zwei (drei) Häufungspunkte beobachtet werden; einer bei etwa 50 % und einer bei etwa 100 % alternativem Allel. **b** Es konnten drei Häufungspunkte beobachtet werden; einer bei etwa 30 %, einer bei etwa 70 % und einer bei etwa 100 % alternativem Allel. **c** Es konnten zwei Häufungspunkte beobachtet werden; einer bei etwa 10 % und einer bei etwa 95 % alternativem Allel. **d** Es konnte ein Häufungspunkt bei etwa 100 % alternativem Allel beobachtet werden.

Mit Hilfe dieser Anforderungen konnte der Clusteralgorithmus „DBSCAN“ als Kandidat für den Nachweis heterogener Populationen beim Mischungsdetektor ausgewählt werden.

5.1.1 DBSCAN-Algorithmus

Der Algorithmus „Density Based Spatial Clustering of Applications with Noise“ (DBSCAN) ist ein dichtebasierter Clusteralgorithmus [52]. Die Definition der Dichte von Objekten ist in Definition 5.1 gegeben.

Definition 5.1 *Dichte von Objekten*

$$\text{Dichte} = \frac{\text{Anzahl Objekte}}{\text{Einnehmender Flächeninhalt}}$$

Das Verfahren basiert auf der Annahme, dass die Dichte von den Datenpunkten innerhalb eines Clusters höher ist, als zwischen unterschiedlichen Clustern. Das Grundprinzip ist in Abbildung 5.2 dargestellt. Bei der Verwendung müssen zwei Parameter (ϵ und *MinPts*) übergeben werden, wobei ϵ die Dichte ist, die innerhalb eines Clusters herrschen soll und *MinPts* die minimale Anzahl an Datenpunkten ist, aus denen ein Cluster bestehen soll. Während der Mustererkennung werden zufällig Datenpunkte ausgewählt und die ϵ -Umgebung dieser Datenpunkte wird auf das Vorhandensein weiterer Datenpunkte untersucht. Befinden sich mehr als *MinPts* Datenpunkte in dieser Umgebung wird dieses Objekt zu einem Kernobjekt des Clusters. Ist dies nicht der Fall, wird dieser Datenpunkt als Rauschen markiert. Eine Anzahl Datenpunkte mit geringer Distanz zueinander bilden eine Menge dichteverbundener Kernobjekte und damit ein Cluster.

Die Wahl geeigneter Werte für die Parameter *MinPts* und ϵ ist von großer Bedeutung für die spätere Qualität der Mustererkennung [52] und sollten daher sorgfältig durchgeführt werden. Ein zu kleines *MinPts* führt zum Beispiel zu vermehrten Clustern aus Rauschobjekten.

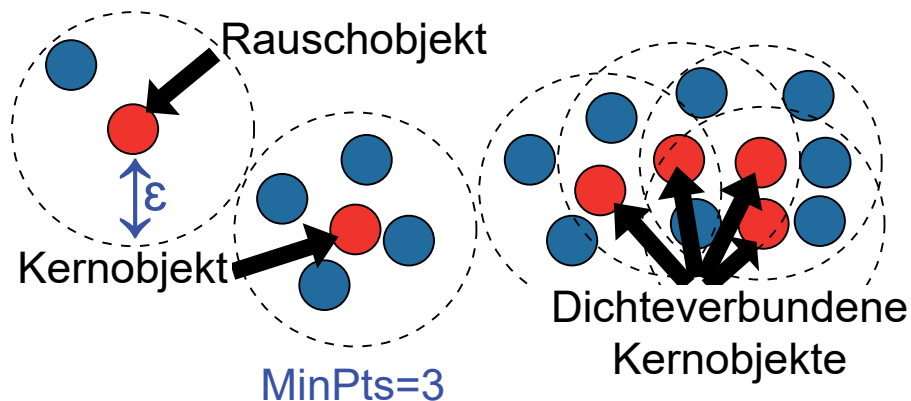


Abbildung 5.2: Schematische Darstellung des DBSCAN-Algorithmus. Befinden sich mehr als $MinPts$ Datenpunkte in der ϵ -Umgebung eines Punktes wird dieser Datenpunkt zu einem Kernobjekt des Clusters. Ist dies nicht der Fall, wird dieser Datenpunkt zum Rauschobjekt. Mehrere Datenpunkte mit geringer Distanz bilden eine Menge von dicht verbundenen Kernobjekten, die zusammen ein Cluster ergeben.

5.1.2 Parametereinstellung des DBSCAN-Algorithmus

Die Parameter des DBSCAN-Algorithmus müssen so eingestellt sein, dass mindestens die drei artifizielle NGS-Proben Mix50, Mix70 und Mix90, korrekt als Mischungen detektiert werden und eine NGS-Probe mit einer reinen Population (Kontrolle) maximal einen Cluster besitzt. Zur Berechnung wurde das R-Paket „fpc“ verwendet. Der Aufruf für den DBSCAN-Algorithmus ist im Quelltext 5.1 aufgeführt.

```
dbscan(x,  $\epsilon$ , MinPts, showplot=1)
```

Quelltext 5.1: Aufruf des DBSCAN-Algorithmus mit R Version 3.0.2.

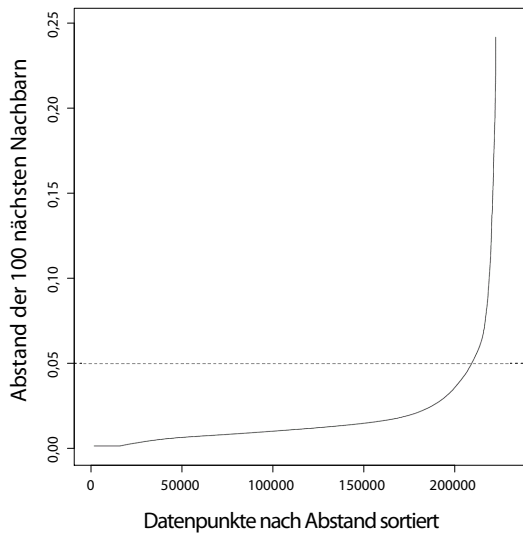
Die Wahl des Parameters $MinPts$ hängt stark von den Daten ab. Die 10% Population der Probe Mix90 war die kleinste zu detektierende Subpopulation in den Testdaten dieser Arbeit. Ungefähr 100 SNPs der NGS-Probe Mix90 hatten eine Frequenz von circa 10% alternativem Allel, sodass dieser Wert als minimale Grenze zur Definition eines Clusters festgelegt wurde. Ein kleinerer Wert würde zu falsch-positiven Clustern führen, beziehungsweise Clustern aus Rauschobjekten. Ein größerer Wert für $MinPts$ würde hingegen dazu führen, dass die SNPs

der 10 % Population der Probe Mix90 als Rauschen klassifiziert werden.

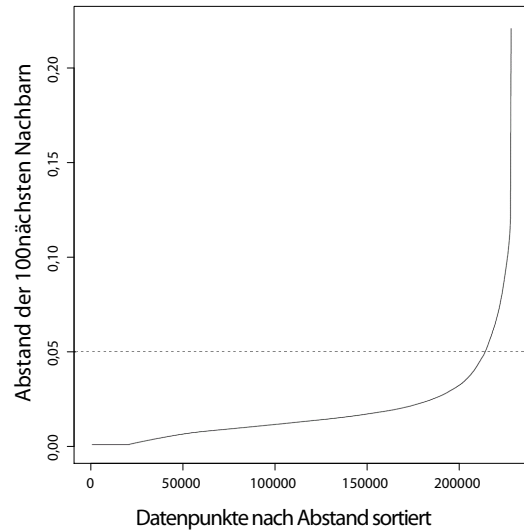
Für die Wahl des ϵ wurden nach Empfehlung von Ester *et al.* [52] k-Abstandsgraf, basierend auf den Abständen der Datenpunkte, erstellt und untersucht, welcher Abstand die Clusterelemente von den Rauschelementen trennt (Abbildung 5.3) [52]. Der Wert k wurde auf $MinPts = 100$ gesetzt und für jeden Datenpunkt der Abstand zu den k-nächsten Nachbarn (kNN) berechnet. Die Datenpunkte wurden anschließend nach dem Abstand der kNN aufsteigend sortiert und abgebildet. Ein „Knick“ in den k-Abstandsgraf zeigt an, welcher Abstand sich für die Wahl von ϵ eignet [52]. Datenpunkte rechts neben diesem Knick können als Rauschobjekte interpretiert werden, während Datenpunkte links von diesem Knick wahrscheinlicher Kernobjekte sind. In Abbildung 5.3 ist zu erkennen, dass für die artifiziellen Mischungen Mix50, Mix70 und Mix90 sowie für die Kontrolle ein Knick bei $\epsilon = 0,05$ zu beobachten war.

Mit den Parametern $MinPts = 100$ und $\epsilon = 0,05$ wurden anschließend die NGS-Proben Mix50, Mix70 und Mix90 sowie die Kontrolle mit Hilfe des DBSCAN-Algorithmus auf Häufungspunkte untersucht. Die Clusterergebnisse sind in Abbildung 5.4 dargestellt. Für die NGS-Probe Mix50 wurden mit dem DBSCAN-Algorithmus die korrekten Häufungspunkte detektiert (grün und rot), die ihre Schwerpunkte bei 48 % und 98 % alternativem Allel hatten (Abbildung 5.4a). Auch für die NGS-Probe Mix70 wurden mit Hilfe des DBSCAN-Algorithmus die drei gesuchten Cluster (blau, rot und grün) mit den Schwerpunkten bei 25 %, 72 % und 98 % alternativem Allel detektiert (Abbildung 5.4b). Das DBSCAN-Clusterergebnis für die NGS-Probe Mix90 zeigte nur zwei statt drei Cluster, da die Dichte zwischen den Datenpunkte der 100 % und der 90 % SNP-Population kleiner ϵ war (Abbildung 5.4c). Für die Kontrolle wurde mit Hilfe des DBSCAN-Algorithmus ein einziger Häufungspunkt mit einem Schwerpunkt bei ungefähr 99 % alternativem Allel detektiert (Abbildung 5.4d).

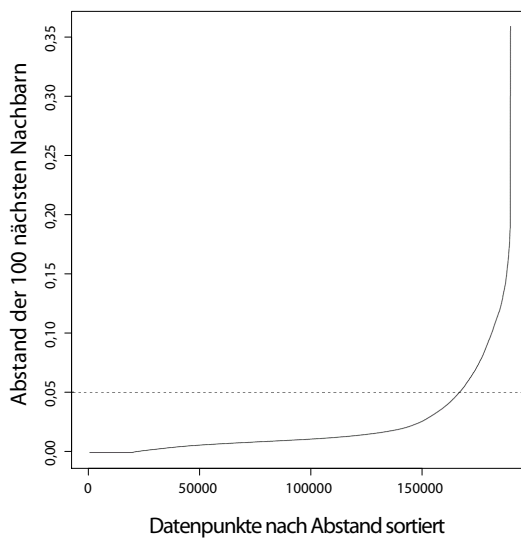
Für die weitere Verwendung des Mischungsdetektors wurde ein Perl-Skript geschrieben, das die BAM-Dateien für die Verwendung mit dem Statistikprogramm R vorbereitet, den Detektor automatisch ausführt und die Clusterergebnisse interpretiert (Quelltext: https://github.com/vschleusener/Skripte_vschleusener). Der erstellte Programmablauf des Mischungsdetektors (Perl-Skript) ist in Abbildung 5.5 schematisch dargestellt. Als Eingabeparameter müssen BAM-Dateien, der Name der Ausgabe-Datei und der Pfad, in dem die Ausgabedateien gespeichert werden sollen, übergeben werden. Zuerst werden die SNPs mittels des GATK-



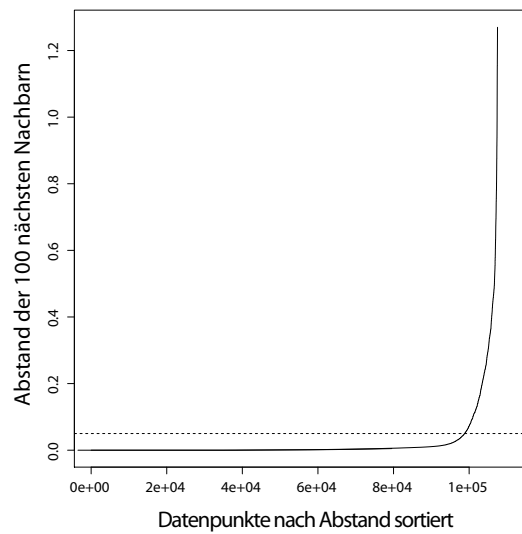
(a) k-Abstandsgraph für die Probe Mix50



(b) k-Abstandsgraph für die Probe Mix70

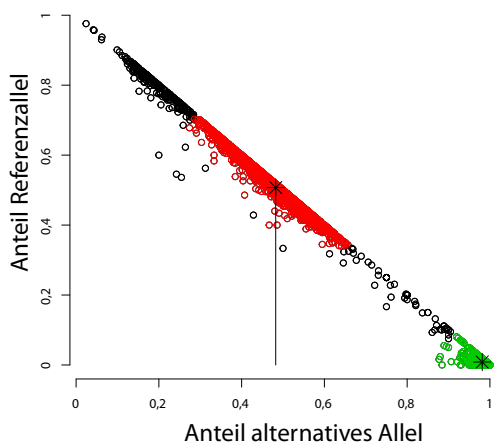


(c) k-Abstandsgraph für die Probe Mix90

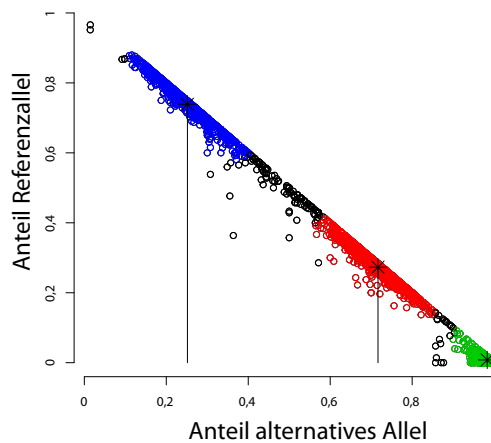


(d) k-Abstandsgraph für die Kontrolle

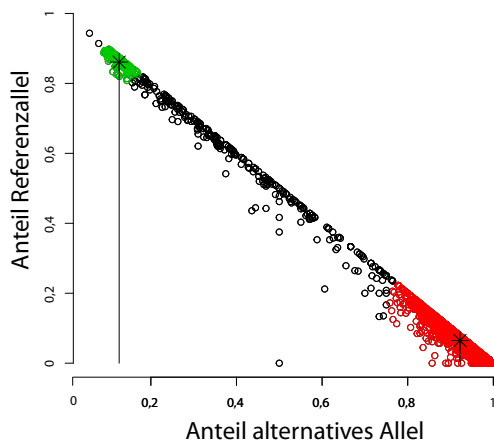
Abbildung 5.3: Sortierte k-Abstandsgrafen für die artifiziiellen Mischungen Mix50, Mix70, Mix90 (a, b, c) und eine Kontrolle (d). Grenzwert für den Abstand ϵ wurde als gestrichelte Linie eingezeichnet. Links von diesem Grenzwert können Objekte als Kernobjekte mit geringer Dichte zueinander interpretiert werden, rechts davon als Rauschobjekte mit großen Abständen zueinander. Berechnet mit R Version 3.1.2



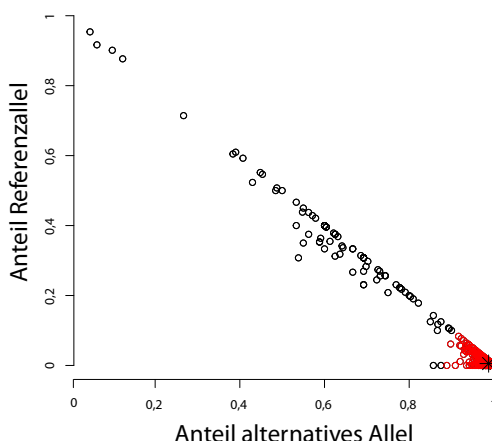
(a) DBSCAN-Clustering für die Probe Mix50



(b) DBSCAN-Clustering für die Probe Mix70



(c) DBSCAN-Clustering für die Probe Mix90



(d) DBSCAN-Clustering für eine Kontrolle

Abbildung 5.4: Ergebnis der Clusteranalyse des DBSCAN-Algorithmus mit $\epsilon = 0,05$ und $MinPts = 100$ für die NGS-Proben Mix50 (a), Mix70 (b), Mix90 (c) und einer Kontrolle (d). a Es wurden zwei Cluster detektiert (rot und grün). b Es wurden drei Cluster detektiert (rot, grün und blau). c Es wurden zwei Cluster detektiert (rot und grün). d Es wurde ein Cluster detektiert (rot).

Algorithmus „UnifiedGenotyper“ anhand der NGS-Daten detektiert. Die SNPs werden aus den erstellten VCF-Dateien in eine Tabelle konvertiert, die mit R eingelesen wird. Mit Hilfe von R werden anschließend die Frequenzen der Referenz- sowie der alternativen Allele berechnet, der DBSCAN-Algorithmus aufgerufen und ein Streudiagramm des Verhältnisses der Referenz- und alternativen Allele erstellt, in dem die detektierten Cluster farblich markiert sind. Als Ausgabe erhält der Nutzer eine Tabelle, die für jede berechnete BAM-Datei die detektierten Häufungsschwerpunkte sowie den interpretierten Status (Mischung ja/nein) enthält. Das Perl-Skript markiert eine Probe als Mischung, wenn von dem DBSCAN-Algorithmus entweder mehr als ein Cluster detektiert wurde oder der Häufungsschwerpunkt bei nur einem detektierten Cluster unterhalb von 0,9 liegt. Für die artifiziellen Mischungen Mix50, Mix70 und Mix90, die als Trainingsdaten verwendet wurden, funktioniert der erstellte Mischungsdetektor sehr gut. Im Weiteren soll getestet werden, wie gut der Detektor auf NGS-Daten klinischer MTBK-Stämme anwendbar ist.

Anpassung des Mischungsdetektors an NextSeq-Daten

Der Mischungsdetektor wurde auf NGS-Datensätzen von MTBK-Stämmen, die mit verschiedenen Illumina-Sequenziergeräten MiSeq, HiSeq 2500 und NextSeq 500 generiert wurden, angewendet. Bei etwa 80 % der NGS-Proben von Illumina NextSeq 500 Geräten detektierte der Mischungsdetektor einen Häufungsschwerpunkt bei ungefähr 20 % alternativem Allel. Bei der detaillierten Analyse dieser Datensätze konnten keine Hinweise auf Mischinfektionen gefunden werden. Zur Überprüfung wurden die Stämme, aus denen die artifiziellen Mischungen Mix50, Mix70 und Mix90 erstellt wurden (Kapitel 2.4), einzeln auf dem NextSeq 500 Gerät sequenziert. Die Ergebnisse sind in Abbildung 5.6 dargestellt.

Für die NextSeq-Daten der Probe 49-02 (ENA-Zugriffsnummer: HG813240) wurde zwar mittels DBSCAN-Algorithmus und den vorab festgelegten Parametern $\epsilon = 0.05$ und $MinPts = 100$ kein zweiter Häufungspunkt bei 20 % alternativem Allel detektiert, dennoch war eine deutliche Häufung in diesem Bereich zu beobachten (Abbildung 5.6a). Für die NextSeq-Daten der Probe 7199-99 (ENA-Zugriffsnummer: HE663067) wurden mit den vorab festgelegten Parametern $\epsilon = 0.05$ und $MinPts = 100$ sogar zwei Häufungspunkte im Bereich von 20 % alternativem Allel detektiert (Abbildung 5.6b).

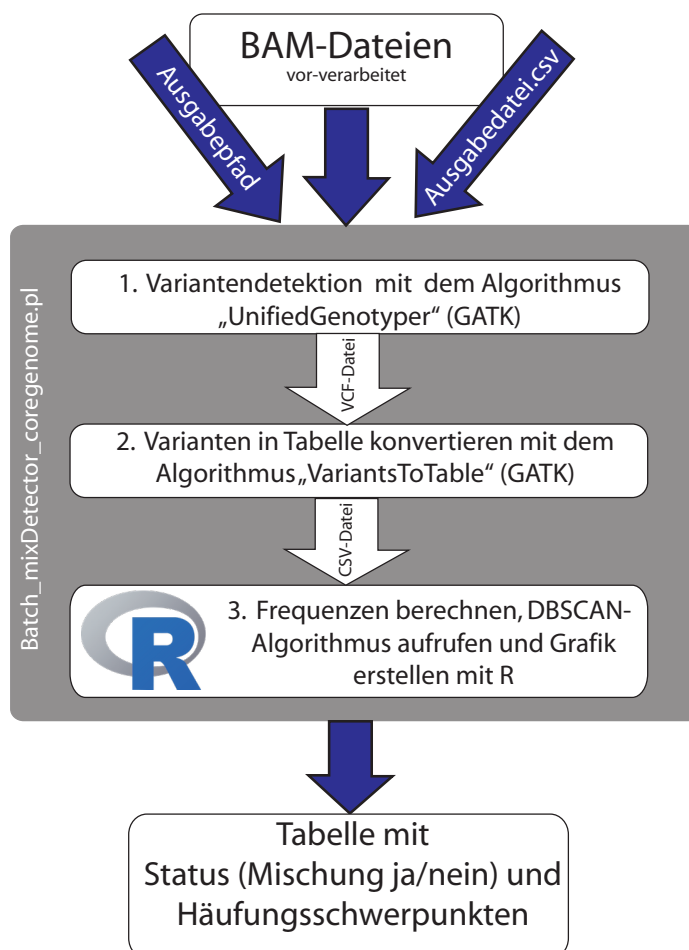
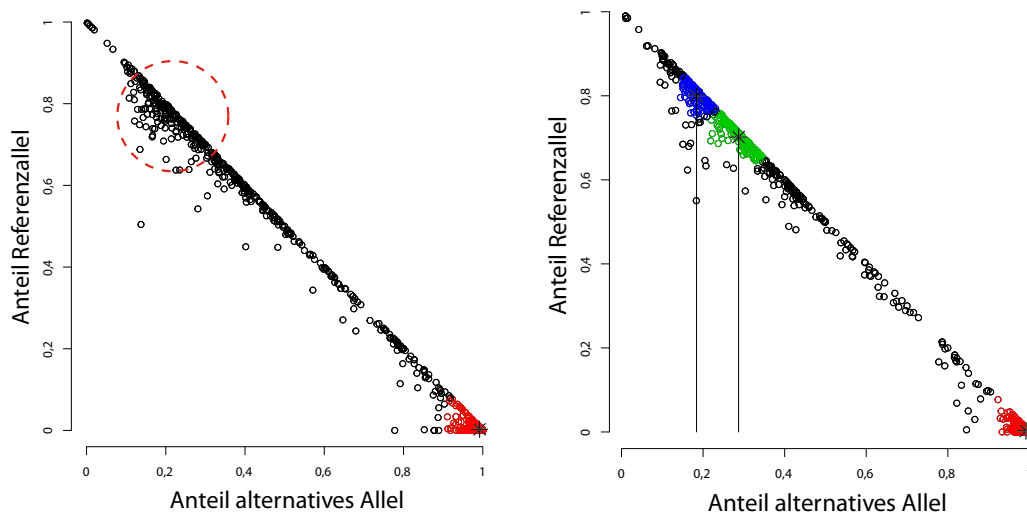


Abbildung 5.5: Schematische Darstellung von `Batch_mixDetector_coregenome.pl` (Quelltext: https://github.com/vschleusener/Skripte_vschleusener). Das Perl-Skript wird im Ordner der zu untersuchenden BAM-Dateien aufgerufen. Die Übergabeparameter sind der Name der Ausgabedatei sowie der Pfad der Ausgabedateien (VCFs, Tabellen und Grafiken). Die Ausgabe ist eine Tabelle mit dem Status (Mischung ja/nein) und den detektierten Häufungsschwerpunkten der untersuchten BAM-Dateien.



(a) DBSCAN-Clustering für die Probe 49-02; $\epsilon = 0.05$ und $MinPts = 100$

(b) DBSCAN-Clustering für die Probe 7199-99; $\epsilon = 0.05$ und $MinPts = 100$

Abbildung 5.6: Ergebnis der Clusteranalyse des DBSCAN-Algorithmus für die NextSeq-Daten der Proben 49-02 (ENA-Zugriffsnummer HG813240, **a**) und 7199-99 (ENA-Zugriffsnummer HE663067, **b**) mit den vorab festgelegten Parametern $\epsilon = 0.05$ und $MinPts = 100$. In **a** wurde ein Cluster detektiert (rot) und in **b** wurden drei Cluster detektiert (rot, grün und blau)

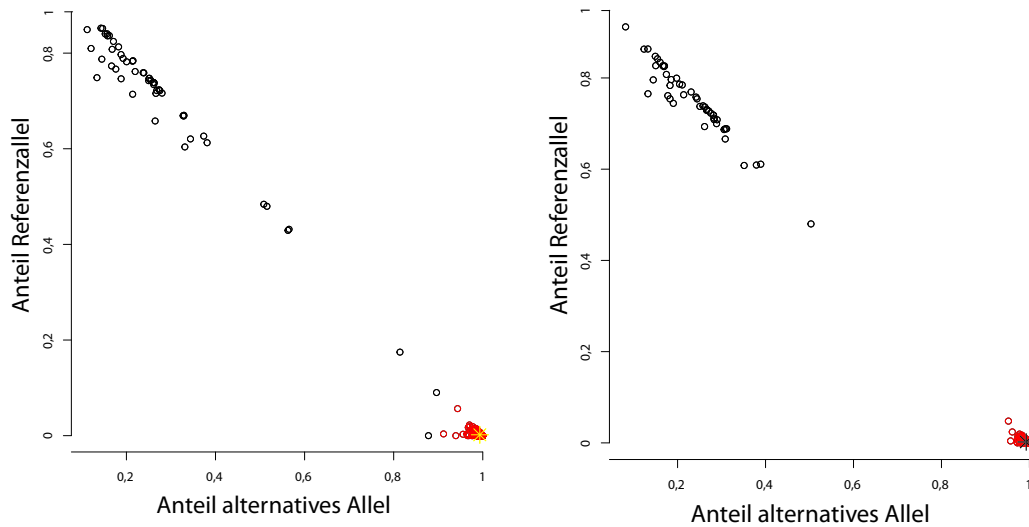
Ein Vergleich der detektierten SNPs mit einer Frequenz unter 75 % alternativem Allel beider NextSeq-Proben zeigte 268 gemeinsame SNPs, die größtenteils in den repetitiv und hoch polymorphen PE- und PPE-Genen detektiert wurden. Aus diesem Grund wurde die SNP-Detektion für den Mischungsdetektor auf das Kerngenom (engl.: core genome) der Studie von Kohl *et al.* aus dem Jahr 2014 beschränkt, das hoch polymorphe und repetitive Gene, wie die PE- und PPE-Gene, ausschließt [108]. Die Clusteranalyse der NextSeq-Daten zeigte anschließend keine Häufungspunkte mehr im Bereich von 20 % alternativem Allel (Daten nicht gezeigt).

Um die Detektierbarkeit der artifiziellen Mischungen Mix50, Mix70 und Mix90 mit dem Mischdetektor zu erhalten, wurden die Parameter des DBSCAN-Algorithmus, wie in Kapitel 5.1.2 beschrieben, mit der Beschränkung auf SNPs des Kerngenoms neu berechnet ($\epsilon = 0,06$ und $MinPts = 55$). In Abbildung 5.7 sind die DBSCAN-Clusterergebnisse für die zwei NextSeq-Daten der beiden Proben 49-02 (Abbildung 5.7a; ENA-Zugriffsnummern: HG813240) und 7199-99 (Abbildung 5.7b; ENA-Zugriffsnummern: HE663067) mit Hilfe der Reduzierung auf SNPs im Kerngenom und der neuen Parameter $\epsilon = 0,06$ und $MinPts = 55$ des DBSCAN-Algorithmus dargestellt. Es wurden keine falschen Häufungspunkte detektiert.

Im Folgenden wird für die Untersuchung von NGS-Proben auf Mischinfektionen der Mischungsdetektor verwendet, der SNPs im Kerngenoms des *M. tuberculosis* detektiert, sowie die Parameter $\epsilon = 0,06$ und $MinPts = 55$ für den DBSCAN-Algorithmus verwendet.

5.1.3 Validierung des Mischungsdetektors

Für die Validierung der Ergebnisse des Mischungsdetektors, wurden NGS-Daten aus Studien von Casali *et al.* aus Russland (ENA-Zugriffsnummer: ERP000192; $n = 1320$) von der ENA-Sequenzdatenbank heruntergeladen [24, 25]. Die verfügbaren FastQ-Dateien wurden mit PhyResSE analysiert, sodass vor-verarbeitete BAM-Dateien erstellt wurden, die anschließend mit Hilfe des implementierten Mischungsdetektors auf Mischungen untersucht wurden. Für die Analyse auf Mischinfektionen, wird eine mittlere Abdeckung von 50 NGS-Sequenzfragmente pro Base vorausgesetzt, beziehungsweise empfohlen ($n = 1206$). Für die NGS-Daten



(a) DBSCAN-Clustering der SNPs des Kerngenoms der Probe 49-02;
 $\epsilon = 0.06$ und $MinPts = 55$

(b) DBSCAN-Clustering der SNPs des Kerngenoms der Probe 7199-99;
 $\epsilon = 0.06$ und $MinPts = 55$

Abbildung 5.7: Ergebnis der Clusteranalyse des DBSCAN-Algorithmus für die NextSeq-Daten der Proben 49-02 (ENA-Zugriffsnummer HG813240, a) und 7199-99 (ENA-Zugriffsnummer HE663067, b) mit den neu berechneten Parametern $\epsilon = 0.06$ und $MinPts = 55$ und der Beschränkung auf SNPs des Kerngenoms. Es wurde jeweils nur ein Cluster (rot) detektiert.

aus der Russland-Studie war der Mischungsstatus (Mischung ja/nein) nicht bekannt, dennoch wurde im Folgenden von einer Mischung ausgegangen, wenn einer der unten genannten Fälle auftrat:

1. Es wurden mindestens drei niederfrequente, phylogenetische SNPs mit ähnlicher Frequenz des alternativen Allels ($> 5\%$) detektiert, die spezifisch für eine von der Hauptpopulation verschiedenen phylogenetische Linie waren; wobei die alternativen Basen mindestens einen Basenqualitätswert von Q20 ($< 1\%$ Fehlerwahrscheinlichkeit, Tabelle 2.1) hatten. Zusätzlich wurden für die phylogenetischen SNPs der Hauptpopulation alternative Allelfrequenzen detektiert, die sich aus dem Abzug des niederfrequenten Anteils von 100% berechnen ließen.
2. Alle detektierten, phylogenetischen SNPs deuteten mit 100% alternativem Allel auf dieselbe phylogenetische Linie hin. Es wurde allerdings mindestens ein resistenzvermittelnder SNP mit einem mittleren Basenqualitätswert der alternativen Allele von mindestens Q30 ($< 0,1\%$ Fehlerwahrscheinlichkeit, Tabelle 2.1) detektiert, bei dem die Frequenz des alternativen Allels mit dem Häufungsschwerpunkt der DBSCAN-Clusteranalyse übereinstimmte.

Um diese Kriterien für alle Proben zu untersuchen, wurde an den spezifischen Positionen der phylogenetischen Linien sowie den resistenzvermittelnden Positionen aus der PhyResSE-Mutationsliste mit Hilfe des Bam-readcount Algorithmus (Kapitel 2.5.2) die NGS-Sequenzfragmente für die jeweiligen Basen gezählt. Anschließend wurden alle 1206 CSV-Dateien (Ausgabe des Perl-Skripts `Batch_Bam-Basecount_pvalue.pl`, Abbildung 5.10) visuell auf die oben genannten Kriterien untersucht und mit den Ergebnissen des Mischungsdetektors verglichen.

In Tabelle 5.1 ist die Vierfeldertafel für den Vergleich des Mischungsdetektors mit der visuellen Untersuchung dargestellt.

Tabelle 5.1: Vierfeldertafel des Mischungsdetektors für die Russland-Studie (ENA Zugriffsnummer: ERP000192).

		Visuelles Ergebnis		Σ
		Mischung	keine Mischung	
Detektor	Mischung	83	37	120
	keine Mischung	0	1086	1086
Σ		83	1123	1206

Der Mischungsdetektor hat alle Proben, die visuell (hier Goldstandard) als Mischinfektionen identifiziert wurden, korrekt als Mischungen detektiert. Darüber hinaus wurden allerdings 37 weitere Proben von dem Mischungsdetektor als Mischinfektionen detektiert, für die keiner der beiden oben genannten Fälle zur visuellen Detektion zu beobachten war. Mit Hilfe dieser Ergebnisse konnte demnach für den Mischungsdetektor eine Sensitivität (richtig-positiv Rate) von 100 % und eine Spezifität (richtig-negativ Rate) von 97 % für die Detektion von Mischinfektionen anhand von NGS-Daten berechnet werden.

Der Großteil der als Mischinfektionen detektierten NGS-Proben (83 %) bestand aus Stämmen mit zwei verschiedenen phylogenetischen Linien. Dennoch konnten 17 Mischinfektionen mit Stämmen der gleichen phylogenetischen Linie und acht Mischinfektionen mit Stämmen aus mehr als zwei verschiedenen phylogenetischen Linien detektiert werden.

In Tabelle 5.2 ist dargestellt, welche phylogenetischen Linien in der Russland-Studie am häufigsten in heterogenen Populationen beobachtet werden konnten. Es ist zu erkennen, dass nicht nur bestimmte phylogenetische Linien koexistieren können (Tabelle 5.2). Am häufigsten (90 %) kamen Mischungen mit einem Stamm der Linie Beijing (Linie 2.2.1) vor, welche allerdings auch die vorherrschende phylogenetische Linie in dieser Population war (65 %, Anhang B, Abbildung B.2). Etwa 25 % der Mischungen waren Mischinfektionen aus Stämmen der Linien Beijing (Linie 2.2.1) und Ural (Linie 4.2.1). Diese Population ist ein Beispiel für eine Region mit vergleichsweise niedriger TB-Inzidenz (84 pro 100 000 Einwohnern) und einer hohen MDR-Rate (20 % der TB-Neuerkrankungen) [212]. In den nächsten Kapiteln werden Mischinfektionen in weitere Populationen mit unterschiedlichen

TB-Inzidenzraten untersucht.

5.1.4 Untersuchung verschiedener Populationen

Im vorherigen Abschnitt konnte gezeigt werden, dass der Mischungsdetektor Mischinfektionen sensitiv detektiert. Im nächsten Abschnitt wurden mit Hilfe des Mischungsdetektors verschiedene Populationen aus unterschiedlichen Regionen der Welt auf das Vorkommen von Mischinfektionen untersucht. Es wurden Regionen mit hoher und niedriger TB-Inzidenz ausgewählt (Abbildung 5.8). In Tabelle 5.3 sind die Ergebnisse der Mischdetektion für die untersuchten Populationen zusammengefasst.

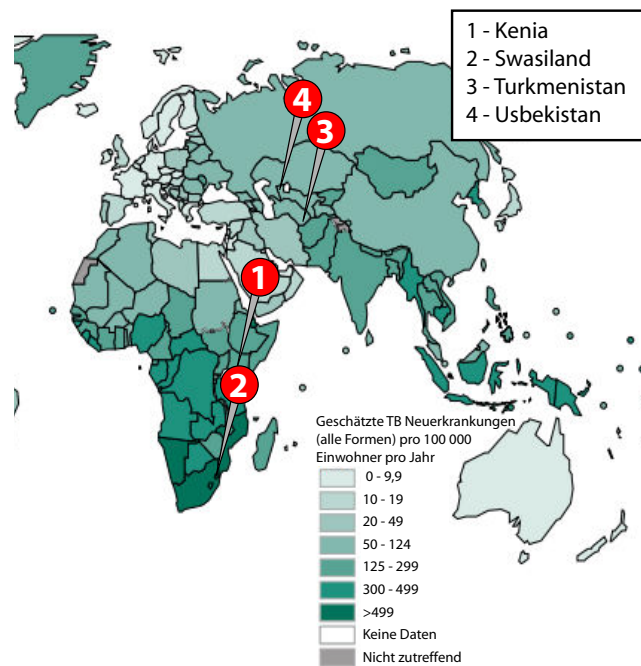


Abbildung 5.8: Inzidenzraten und Lage der untersuchten Regionen.

Der größte Anteil von Mischinfektionen in der Stichprobe wurde mit fast 10 % in der untersuchten Population aus Turkmenistan detektiert (Tabelle 5.3).

Tabelle 5.2: Vorkommen der verschiedenen phylogenetischen Linien in den heterogenen NGS-Proben der Russland-Population.

Mischungen	Anzahl
Beijing-Beijing	14
Beijing-Ural	16
Beijing-LAM	16
Beijing-Haarlem	10
Beijing-4.8	10
Beijing-4.1	2
Beijing-EAI	1
Beijing-S-type	1
4.8-4.8	3
4.8-4.1	1
4.8-Haarlem	2
4.8-LAM	1
4.1-LAM	1
Ural-4.8	3
LAM-WA2	1
Haarlem-Ural	1

Tabelle 5.3: Ergebnisse des Mischungsdetektors für die untersuchten Populationen.

ID	Land	# Genome	#Misch- ungen	Anteil
1	Kenia	226	20	0,088
2	Swasiland	280	7	0,025
3	Turkmenistan	71	7	0,099
4	Usbekistan	305	6	0,020

Mischungsdetektion in einer Population aus Kenia

Mit insgesamt knapp 90 000 TB-Fällen im Jahr 2015 ist Kenia ein Land mit hohen TB-Raten [213]. Die Studienpopulation besteht aus 226 Proben von erwachsenen Patienten mit einer neu-diagnostizierten TB, die in verschiedenen Kliniken (Rhodes city council Klinik, Aga Khan Krankenhaus, Dagoretti und Waithaka Kliniken in Nairobi) rekrutiert wurden. Mit Hilfe des Mischungsdetektors wurden 20 Mischungen identifiziert. Die Ergebnisse der Mischdetektion für die Population aus Kenia sind in Tabelle 5.4 dargestellt. Der Großteil mit 80 % der Mischungen bestand aus Mischinfektionen zweier Stämme mit unterschiedlichen phylogenetischen Linien. In der Probe 13 konnten sogar drei Stämme mit verschiedenen phylogenetischen Linien (Linie 4.1.2.1, Linie 4.3.4.2 und Linie 3.1.1) detektiert werden, die jeweils ungefähr 30 % der Gesamtpopulation ausmachten. Bei den NGS-Proben 18, 19 und 20 handelte es sich um Mischungen von Stämmen derselben phylogenetischen Linie, wobei nur für die Probe 18 der Häufungsschwerpunkt durch die alternative Allelfrequenz der resistenzvermittelnden SNPs bestätigt wurde. Für alle NGS-Proben bis auf die NGS-Proben 16, 18, 19 und 20 passte die Frequenz der linienspezifischen SNPs zu den detektierten Häufungsschwerpunkten des Mischungsdetektors (Tabelle 5.4). Bei fünf der 20 detektierten Mischinfektionen (NGS-Proben 2, 13, 15, 16 und 18) stimmte außerdem die Frequenz des alternativen Allels der resistenzvermittelnden SNPs mit dem detektierten Häufungsschwerpunkt des Mischungsdetektors überein (Tabelle 5.4).

Tabelle 5.4: Mischdetektion in einer Population aus Kenia.

ID	Cluster 1	Cluster 2	Cluster 3	Linie 1	Freq	Linie 2	Freq	Linie 3	Freq	Resistenz
1	0,9934535	0,5004947		Beijing/2.2.1	0,5	TUR/4.2.2	0,5			
2	0,9933762	0,8677914	0,1339923	LAM/4.3.3	0,9	Beijing/2.2.1	0,15			INH,RMP,SM ^a
3	0,9491556	0,1153508		4.8	0,9	S-type/4.4.1.1	0,1			
4	0,9939560	0,6458172	0,3360273	Beijing/2.2.1	0,35	TUR/4.2.2	0,66			
5	0,9940000	0,8212106	0,1733934	Haarlem/4.1.2.1	0,8	Uganda/4.6.1.1	0,17			
6	0,9966218	0,6941897	0,2985986	Beijing/2.2.1	0,7	S-type/4.4.1.1	0,3			
7	0,9944114	0,7354382	0,2530784	LAM/4.3.3	0,75	4.1.1.1/Gh,Ha,X ^b	0,25			
8	0,9261046	0,1366900		EAI/1.1.2	0,9	Beijing/2.2.1	0,1			
9	0,9925923	0,4776777		LAM/4.3.4.2.1	0,5	Delhi/CAS/3.1.1	0,57			
10	0,9957400	0,4519013		Haarlem/4.1.2.1	0,55	Delhi/CAS/3.1.1	0,49			
11	0,9929225	0,7822792	0,2142441	Delhi/CAS/3.1.1	0,75	4.1.1.1/Gh,Ha,X ^b	0,25			
12	0,9940679	0,6201714	0,3709882	Haarlem/4.1.2.1	0,57	Delhi/CAS/3.1.1	0,35			
13	0,9936512	0,3156102		Haarlem/4.1.2.1	0,3	LAM/4.3.4.2	0,3	Delhi/CAS/3	0,3	INH ^c
14	0,9938951	0,8325838	0,1703223	Haarlem/4.1.2.1	0,8	Beijing/2.2.1	0,15			INH,RMP,SM,EMB,PZA ^d
15	0,8846139	0,1532173		EAI/1.1.2	0,85	Haarlem/4.1.2.1	0,15			RMP ^e
16	0,9867646	0,1374500		4.8	0,95	<i>M. canettii</i>	0,05			
17	0,9962922	0,6740100	0,3213232	S-type/4.4.1.1	0,68	Delhi/CAS/3.1.1	0,35			
18	0,9922639	0,1539061		Delhi/CAS/3.1.1	1					
19	0,9939735	0,1372086		LAM/4.3.3	1					RMP,AMK/KAN/CPM ^f
20	0,9928451	0,1416488		LAM/4.3.4.1	1					

Abkürzungen: AMK-Amikacin, CPM-Capreomycin, EMB-Ethambutol, Freq-Frequenz, INH-Isoniazid, KAN-Kanamycin, PZA-Pyrazinamid, RMP-Rifampicin, SM-Streptomycin.

^a Alle SNPs mit 80 % alternativem Allel

^b Eine der Linie Ghana, Haarlem oder X-type. Nicht weiter differenzierbar.

^c SNP hat 40 % alternatives Allel

^d INH-SNP in 100 % vorhanden, die anderen mit ungefähr 85 % alternativem Allel

^e RMP-SNP mit 5 % alternativem Allel

^f Beide SNPs mit etwa 20 % alternativem Allel

Mischungsdetektion in einer Population aus Swasiland

Die TB-Inzidenz in Swasiland ist mit 1 382 Fällen pro 100 000 Einwohnern im Jahr 2015 im Vergleich sehr hoch. Die 280 Proben der Studienpopulation stammen aus einer Querschnittsstudie zur Resistenzüberwachung, die zwischen 2009 und 2010 in Swasiland durchgeführt wurde [174]. Es wurden Patienten aus 15 Studienzentern mit neu diagnostizierter TB und wiederkehrender, bereits behandelter TB eingeschlossen. Die detektierten Mischinfektionen des Mischungsdetektors sind in Tabelle 5.5 dargestellt. Bei allen NGS-Proben, die vom Mischungsdetektor als Mischungen detektiert wurden, konnten bis auf eine Ausnahme (Probe 1) jeweils linienspezifische SNPs detektiert werden, deren alternative Allelfrequenz mit den detektierten Häufungsschwerpunkten übereinstimmte (Tabelle 5.5). Eine Ausnahme bildete die NGS-Probe 1, bei der sich weder über linienspezifische SNPs, noch über die Frequenz der alternativen Allele der resistenzvermittelnden SNPs Hinweise auf eine Mischinfektion finden ließen. Alle anderen NGS-Proben bestanden aus mindestens zwei Stämmen verschiedener phylogenetischer Linien. Bei der Probe 4 wurden sogar spezifische SNPs dreier verschiedener phylogenetischer Linien detektiert (Linie 2.2.1, Linie 4.3.2.1 und Linie 4.1.1.2; Tabelle 5.5), allerdings war die Frequenz des alternativen Allels der spezifischen SNPs der Linie 4.1.1.2 mit 2% sehr gering. Bei der Probe 5 handelte es sich anhand der Nomenklatur nach dem Ursprungsland um dieselbe phylogenetische Linie (LAM), die phylogenetische Klassifizierung nach Coll *et al.* [32] erlaubte allerdings eine weitere Unterscheidung in zwei Sublinien dieser phylogenetischen Linie (Linie 4.3.2.1 und Linie 4.3.4.2.1; Tabelle 5.5).

Zusätzlich stimmte die Frequenz des alternativen Allels der resistenzvermittelnden SNPs der NGS-Proben 2, 3, 4 und 5 mit einem der detektierten Häufungsschwerpunkten des Mischungsdetektors überein und konnte somit einem Stamm der jeweiligen Mischinfektion zugeordnet werden (Tabelle 5.5).

Tabelle 5.5: Stämme mit einer Mischinfektion aus einer Population aus Swasiland.

ID	Cluster 1	Cluster 2	Cluster 3	Linie 1	Freq	Linie 2	Freq	Linie 3	Freq	Resistenz
1	0,9940259	0,2380786		Beijing/2.2.1	1					INH,RMP,SM,EMB,PZA ^a
2	0,9485848	0,1294910		Haarlem/4.1.2.1	0,9	LAM/4.3.4.2.1	0,1			INH,RMP,SM,PZA ^b
3	0,9384353	0,1281049		S-type/4.4.1.1	0,85	Beijing/2.2.1	0,15			INH,RMP,SM,EMB ^b
4	0,9957806	0,7835512	0,1994432	Beijing/2.2.1	0,85	LAM/4.3.2.1	0,16	4.1.1.2/G,H,X ^c	0,02	SM ^d
5	0,9958997	0,6186731	0,3746979	LAM/4.3.2.1	0,6	LAM/4.3.4.2.1	0,4			INH,SM ^e
6	0,9907974	0,8139996	0,1870446	S-type/4.4.1.1	0,82	LAM/4.3.4.2.1	0,17			
7	0,9153576	0,1331659		Beijing/2.2.1	0,83	G,H,X/4.1.1.3 ³	0,15			

Abkürzungen: EMB-Ethambutol, Freq-Frequenz, INH-Isoniazid, PZA-Pyrazinamid, RMP-Rifampicin, SM-Streptomycin.

^a Alle SNPs mit 100 % alternativem Allel

^b Alle SNPs mit ungefähr 90 % alternativem Allel

^c Aufgrund der niedrigen Frequenz unsichere Vorhersage. Eine der Linie Ghana, Haarlem oder X-type. Nicht weiter differenzierbar.

^d SM-SNP mit ungefähr 20 % alternativem Allel

^e INH-SNP mit 35 % alternativem Allel und SM-SNP mit ungefähr 65 % alternativem Allel

Mischungsdetektion in einer Population aus Turkmenistan

Turkmenistan hatte eine TB-Inzidenz von 72 Fällen pro 100 000 Einwohnern im Jahr 2015 und ist damit vergleichsweise gering (Abbildung 5.8). Die 71 NGS-Proben der Studienpopulation sind größtenteils MDR-TB Fälle und stammen aus einer Studie zur Überwachung der von der WHO empfohlenen DOTS-Strategie zur Eindämmung der TB-Infektionen. Die sieben mit dem Mischungsdetektor detektierten Mischinfektionen aus Turkmenistan sind in Tabelle 5.6 dargestellt. Bei sechs der sieben Mischungen konnten jeweils spezifische SNPs für zwei verschiedene phylogenetische Linien detektiert werden, deren alternative Allelfrequenz mit dem detektierten Häufungsschwerpunkt des Mischungsdetektors übereinstimmten (Tabelle 5.6). In der Probe 3 konnten sogar spezifische SNPs für drei verschiedene phylogenetische Linien detektiert werden (Linie 4.1.2, 60 %; Linie 4.4.1.1, 35 % und Linie 2.2.1, 5 %; Tabelle 5.6). Insgesamt war an allen Mischinfektionen ein Stamm der Linie Beijing (Linie 2.2.1) beteiligt (Tabelle 5.6).

Zusätzlich stimmte die alternative Allelfrequenz der detektierten resistenzvermittelnden SNPs bis auf eine Ausnahme mit jeweils einem Häufungsschwerpunkt überein, sodass diese jeweils einem der beteiligten Stämme der Mischinfektion zugeordnet werden konnten (Tabelle 5.6). Bei der Probe 3 war die Zuordnung nicht eindeutig, da beide Stämme in 50 % vorkamen. Es wurden für diese NGS-Probe außerdem drei verschiedene RMP-resistenzvermittelnde SNPs mit Frequenzen jeweils unter 20 % alternativem Allel detektiert (Tabelle 5.6).

Tabelle 5.6: Mischdetektion in einer Population aus Turkmenistan.

ID	Cluster 1	Cluster 2	Cluster 3	Linie 1	Freq	Linie 2	Freq	Linie 3	Freq	Resistenz
1	0,9956193	0,7932146	0,2052175	Ural/4.2.1	0,8	Beijing/2.2.1	0,2			INH,RMP,SM,EMB,PZA ^a
2	0,9479945	0,1216887		Haarlem/4.1.2	0,9	Beijing/2.2.1	0,1			INH,RMP,SM,EMB,PZA ^b
3	0,9946428	0,5069957		LAM/4.3.3	0,5	Beijing/2.2.1	0,5			INH,RMP,SM,EMB ^c
4	0,9434367	0,1129735		Linie 4.5	0,9	LAM/4.3.3	0,1			
5	0,9806619	0,7626516	0,2252129	Beijing/2.2.1	0,76	Linie 4.8	0,23			INH,RMP,SM,PZA ^d
6	0,9940224	0,6138258	0,3212697	Haarlem/4.1.2	0,6	S-type/4.4.1.1	0,35	Beijing/2.2.1	0,05	INH,RMP,SM,EMB ^e
7	0,9425242	0,1205731		Ural/4.2.1	0,9	Beijing/2.2.1	0,1			INH,RMP,SM ^f

Abkürzungen: EMB-Ethambutol, Freq-Frequenz, INH-Isomiazid, KAN-Kanamycin, PZA-Pyrazinamid, RMP-Rifampicin, SM-Streptomycin.

^a Jeweils ein Resistenz-SNP für INH, RMP, SM und EMB in 80 % sowie ein SM- und PZA-resistenzvermittelnder SNP mit ungefähr 20 % alternativem Allel

^b Alle SNPs mit ungefähr 10 % alternativem Allel

^c INH-, SM- und EMB-resistenzvermittelnder SNP mit ungefähr 50 % alternativem Allel. Drei RMP-resistenzvermittelnde SNPs mit 17 %, 10 % und 14 % sowie ein SNP für AMK/KAN/CPM-Resistenz mit 35 % alternativem Allel

^d Alle SNPs mit ungefähr 80 % alternativem Allel

^e Alle SNPs haben eine Frequenz unter 10 % alternativem Allel

^f INH-resistenzvermittelnder SNP mit 100 % alternativem Allel. RMP- und SM-resistenzvermittelnder SNP mit 10 % alternativem Allel

Mischungsdetektion in einer Population aus Usbekistan

Die TB-Inzidenz ist mit 80 pro 100 000 Einwohnern im Jahr 2015 vergleichsweise gering (Abbildung 5.8), allerdings ist die MDR-TB Rate in Usbekistan mit über 20 % der TB-Neuerkrankungen sehr hoch (Abbildung 1.2). Bei der verwendeten Stichproben handelte es sich um MDR-TB Fälle aus der Region Nukus ($n = 305$). Mit Hilfe des Mischungsdetektors wurden sechs Mischungen detektiert (Tabelle 5.7). Alle detektierten Mischungen enthielten einen Stamm der Linie Beijing (Linie 2.2.1). Bei der Hälfte der Mischungen ($n = 3$) wurden neben den beijingspezifischen SNPs keine spezifischen SNPs anderer phylogenetischer Linien detektiert (Proben 1, Probe 4 und Probe 6; Tabelle 5.7) und auch die resistenzvermittelnden SNPs der Proben 1 und 4 wurden alle mit 100 % alternativem Allel detektiert (Tabelle 5.7). Bei der Probe 6 stimmten allerdings die alternativen Allelfrequenzen der resistenzvermittelnden SNPs mit den detektierten Häufungsschwerpunkten des Mischungsdetektors überein (Tabelle 5.7). Ein ähnliches Ergebnis konnte für alle anderen detektierten Mischinfektionen beobachtet werden. Es wurde bei jeder Probe mindestens ein resistenzvermittelnder SNP detektiert, der mit dem Häufungsschwerpunkt des Mischungsdetektors übereinstimmte, sodass eine Zuordnung der resistenzvermittelnden SNPs zu jeweils einem beteiligten Stamm der Mischinfektion möglich war (Tabelle 5.7). Eine Ausnahme bildete die Probe 2, bei der sich die resistenzvermittelnden SNPs nicht anhand der Frequenz des alternativen Allels einem Stamm der Mischinfektion zuordnen ließen, da alle SNPs mit ungefähr 50 % alternativem Allel vorkamen (Tabelle 5.7). Außerdem wurden für Probe 2 zwei FQ-resistenzvermittelnde SNPs mit 15 % und 38 % alternativem Allel nachgewiesen (Tabelle 5.7).

Tabelle 5.7: Mischdetektion in einer Population aus Usbekistan (Nukus).

ID	Cluster 1	Cluster 2	Cluster 3	Linie 1	Freq	Linie 2	Freq	Linie 3	Freq	Resistenz
1	0,9920093	0,1436487		Beijing/2.2.1	1					INH,RMP,SM,EMB,PZA ^a
2	0,9917961	0,4819502		Beijing/2.2.1	0,5	Haarlem/4.1.2.1	0,5			INH,RMP,SM,EMB,PZA,FQ ^b
3	0,9794689	0,7377587	0,2343093	Haarlem/4.1.2.1	0,8	Beijing/2.2.1	0,2			INH,RMP,SM,EMB,PZA ^c
4	0,9944447	0,2122428		Beijing/2.2.1	1					INH,RMP,SM,EMB,PZA ^a
5	0,9914082	0,7979720	0,1958258	Beijing/2.2.1	0,85	Linie 4-8	0,15			INH,RMP,SM,EMB,PZA, AMK/KAN/CPM,FQ ^d
6	0,9890861	0,8188997	0,1643307	Beijing/2.2.1	1					INH,RMP,SM,EMB,PZA ^e

Abkürzungen: AMK-Amikacin, CPM-Capreomycin, EMB-Ethambutol, Freq-Frequenz, FQ-Fluorochinolone, INH-Isoniazid, KAN-Kanamycin, PZA-Pyrazinamid, RMP-Rifampicin, SM-Streptomycin.

^a Alle SNPs in 100 % vorhanden

^b INH- und RMP-resistenzvermittelnder SNP in 100 % vorhanden; SM-, PZA- und EMB-resistenzvermittelnder SNP mit etwa 50 % alternativem Allel vorhanden und zwei FQ-resistenzvermittelnde SNPs mit 15 % bzw. 38 % alternativem Allel

^c INH- und RMP-resistenzvermittelnder SNPs mit jeweils zwei verschiedenen Allelen an gleicher Position mit ungefähr 20 % bzw. 80 % alternativem Allel. Zwei SM-resistenzvermittelnder SNPs mit jeweils 25 % alternativem Allel. PZA- und EMB-resistenzvermittelnder SNP mit jeweils 30 % alternativem Allel

^d RMP- und EMB-resistenzvermittelnder SNP in 100 % vorhanden. Jeweils zwei SNPs für INH- und PZA-Resistenz mit 15 % bzw. 80 % alternativem Allel. FQ- und AMK/KAN/CPM-resistenzvermittelnder SNP mit jeweils ungefähr 20 % alternativem Allel

^e SM- und EMB-resistenzvermittelnder SNP in 100 % vorhanden. Jeweils zwei SNPs für INH- und RMP-Resistenz mit 15 % bzw. 85 % alternativem Allel. PZA-resistenzvermittelnder SNP mit 80 % alternativem Allel.

5.2 Heteroresistenzen

Zurzeit fehlen für die molekularbiologische TB-Resistenzdiagnostik basierend auf NGS-Genomdaten von MTBK-Stämmen valide Algorithmen zur Bestimmung niederfrequenter Mutationen in Resistenzgenen. Ein zusätzliches Problem stellt die Bewertung dieser Mutationen im Hinblick auf eine, für die Therapie relevante, Resistenzausbildung dar. Aus diesem Grund wird in den folgenden Abschnitten die Entwicklung eines Ansatzes zur Detektion und Bewertung von niederfrequenten SNPs anhand von Genomsequenzdaten klinischer MTBK-Stämme beschrieben. Mit Hilfe der etablierten Methodik, werden die Möglichkeiten zum Nachweis niederfrequenter Resistenzmutationen basierend auf der Analyse artifizierlicher Mischungen von sensiblen und resistenten MTBK-Stämmen und von NGS-Daten klinischer Isolate evaluiert.

5.2.1 Binomialansatz

Eine Hauptproblematik bei der Bestimmung niederfrequenter Mutationen stellt die Unterscheidung tatsächlicher, in der Bakterienpopulation vorliegender Mutationen von Sequenzierfehlern dar. Die heutigen NGS-Sequenziergeräte produzieren Sequenzierfehler nach dem Zufallsprinzip [115]. Dabei muss zwischen zwei Arten von Fehlern unterschieden werden, echten Sequenzierfehlern und schlecht-detektierten Basen. Als Sequenzierfehler werden im Folgenden mutierte Basen bezeichnet, die von dem Sequenziergerät einen guten Basenqualitätswert zugeordnet bekommen haben und dennoch nicht in der biologischen Probe vorhanden sind. Die Firma Illumina gibt für die eigenen Geräte eine Wahrscheinlichkeit von 0,01 % für diese Art der Fehler an. Weiterhin gibt es bei jedem Sequenzierlauf Basen mit einem geringen Basenqualitätswert Q (Gleichung 2.1). Generell lässt sich jeder Basenqualitätswert in die Fehlerwahrscheinlichkeit transformieren (Gleichung 5.1).

$$e = P(\text{Fehler}) = 10^{-\frac{Q}{10}} \quad (5.1)$$

Es wurden mit den Selektionsklonen SR4k (*rpoB* 526 Mutation; Genomposition 761140 A → C) und SR1a (*rpoB* 531 Mutation; Genomposition 761155 C → T) jeweils drei Mischungen mit 1 %, 5 % und 10 % Selektionsklon und dem Laborstamm *M. tuberculosis* H37Rv ATCC 27294 erstellt und sequenziert (Tabelle 2.3).

Zur Berechnung einer mittleren Fehlerwahrscheinlichkeit an den resistenzvermittelnden Positionen wurden die Basenqualitätswerte der einzelnen Basen der artifizialen NGS-Mischproben (SR4k_rpoB526_1, SR4k_rpoB526_5, SR4k_rpoB526_10, SR1a_rpoB531_1, SR1a_rpoB531_5 und SR1a_rpoB531_10, Tabelle 2.3) mit Hilfe der Gleichung 5.1 in die Fehlerwahrscheinlichkeit transformiert. Anschließend wurden diese für jede Position gemittelt und von den Positionsmittelwerten der Durchschnitt gebildet (Gleichung 5.2).

$$M(FW_{gesamt}) = \frac{M(FW_{Pos1}) + M(FW_{Pos2}) + \dots + M(FW_{Pos369})}{\# \text{ Pos (369)}} \quad (5.2)$$

M := Mittelwert; FW := Fehlerwahrscheinlichkeit

Für die artifizial hergestellten, heteroresistenten NGS-Proben (Tabelle 2.3) lag die Fehlerwahrscheinlichkeit basierend auf den Basenqualitätswerten an den resistenzvermittelnden Positionen demnach im Mittel bei 1,2%. Die Sequenzierfehler, ebenso wie Basen schlechter Qualität, sind pro Position unabhängig über die ausgerichteten NGS-Sequenzfragmente verteilt. Aus diesem Grund konnte ihre Verteilung mit Hilfe der Binomialverteilung angenähert werden.

5.2.2 Binomialverteilung

Die Binomialverteilung ist eine diskrete Wahrscheinlichkeitsverteilung mit zwei Parametern (n, p) [67]. Die Verteilung leitet sich aus der Verkettung von Bernoulli-Experimenten zu einer Bernoulli-Kette her [15, 67]. Die einzelnen Zufallsexperimente der Kette müssen unabhängig voneinander mit gleichbleibender Erfolgswahrscheinlichkeit sein und dürfen lediglich zwei mögliche Ausgänge (Erfolg oder Misserfolg; hier: Referenzbase oder Nicht-Referenzbase) haben [15, 67]. Die Binomialverteilung beschreibt die Wahrscheinlichkeit für das Auftreten einer bestimmten Anzahl von Erfolgen (z. B.: Anzahl der Nicht-Referenzbasen an einer Position). Die Wahrscheinlichkeitsfunktion ist in Definition 5.2 gegeben.

Definition 5.2 *Wahrscheinlichkeitsfunktion der Binomialverteilung.*

Sei p die Erfolgswahrscheinlichkeit und n die Anzahl Experimente, dann gilt:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, x \geq 0$$

Der Erwartungswert berechnet sich als $\mu = np$ und die Varianz als $\sigma = \sqrt{np(1-p)}$ [67].

Der Binomialtest ist ein statistischer Test, bei dem die Teststatistik binomialverteilt ist. Er kann verwendet werden, um Hypothesen über dichotome Merkmale, Merkmale mit zwei Ausprägungen, zu prüfen [141]. Eine typische Fragestellung des Binomialtests ist: „Unterscheidet sich die beobachtete Auftretenswahrscheinlichkeit von der theoretisch erwarteten Auftretenswahrscheinlichkeit [141]?“

Die Hypothesen des Binomialtests für die Problemstellung der vorliegenden Arbeit lauten:

Nullhypothese: $H_0 : p \leq p_0$

Alternativhypothese: $H_1 : p > p_0$

p = Wahrscheinlichkeit für einen Fehler in den Daten;

Anteil nicht übereinstimmender Basen

p_0 = Generelle Wahrscheinlichkeit für Sequenzierfehler

+ mittlere Fehlerwahrscheinlichkeit anhand der Basenqualitätswerte

Es soll demnach überprüft werden, ob an einer bestimmten Genomposition in den ausgerichteten NGS-Sequenzfragmenten signifikant mehr nicht-übereinstimmende Basen beobachtet werden können, als unter der Annahme der Nullhypothese (Fehlerwahrscheinlichkeit anhand der Basenqualitätswerte + Wahrscheinlichkeit für Sequenzierfehler) erwartet werden. Das Signifikanzniveau für den Binomialtest wurde auf $\alpha = 0,05$ gesetzt.

Ein weiteres Kriterium zur Detektion niederfrequenter Mutationen war die Anzahl an NGS-Sequenzfragmenten, die eine Position mindestens abdecken müssen, um einen bestimmten Anteil nicht-übereinstimmender Basen detektieren zu können. Aus diesem Grund wird in dem nächsten Abschnitt eine Simulation beschrieben, die einen Anhaltspunkt liefern konnte.

5.2.3 Abdeckungsuntersuchung

Für die Simulation der mit dem Binomialtest minimal-detektierbaren Frequenz des alternativen Allels in Abhängigkeit der benötigten Mindestabdeckung wurde eine Fehlerwahrscheinlichkeit von 1,3% gewählt, die sich aus der mittleren Fehlerwahrscheinlichkeit der heteroresistenten NGS-Proben (1,2%) und der Illumina Fehlerrate (0,1%) [69] zusammensetzt. Die Abdeckung variierte in diesem Test zwischen 50 und 1000 NGS-Sequenzfragmenten pro Position (Abbildung 5.9). Die Abbildung zeigt abhängig von der Abdeckung die mit dem Binomialtest minimal detektierbare Frequenz der mutierten Base. Anhand dieser Simulation liegt die Grenze für detektierbare Subpopulationen bei etwa 2%, sofern eine Abdeckung bis zu 1000 NGS-Sequenzfragmenten erzielt werden kann. Eine Frequenz von 2,5% alternativem Allel konnte bereits ab einer mittleren Abdeckung von 360 NGS-Sequenzfragmenten detektiert werden. Verringert sich die Abdeckung, so erhöht sich die Detektionsschwelle (Abbildung 5.9).

Die artifiziellen, heteroresistenten NGS-Proben wurden mit einer mittleren Abdeckung von 600 NGS-Sequenzfragmenten sequenziert (siehe Kapitel 2.4). Im nächsten Abschnitt werden die Selektionsklonmischungen (Tabelle 2.3) mit Hilfe des Binomialansatz analysiert.

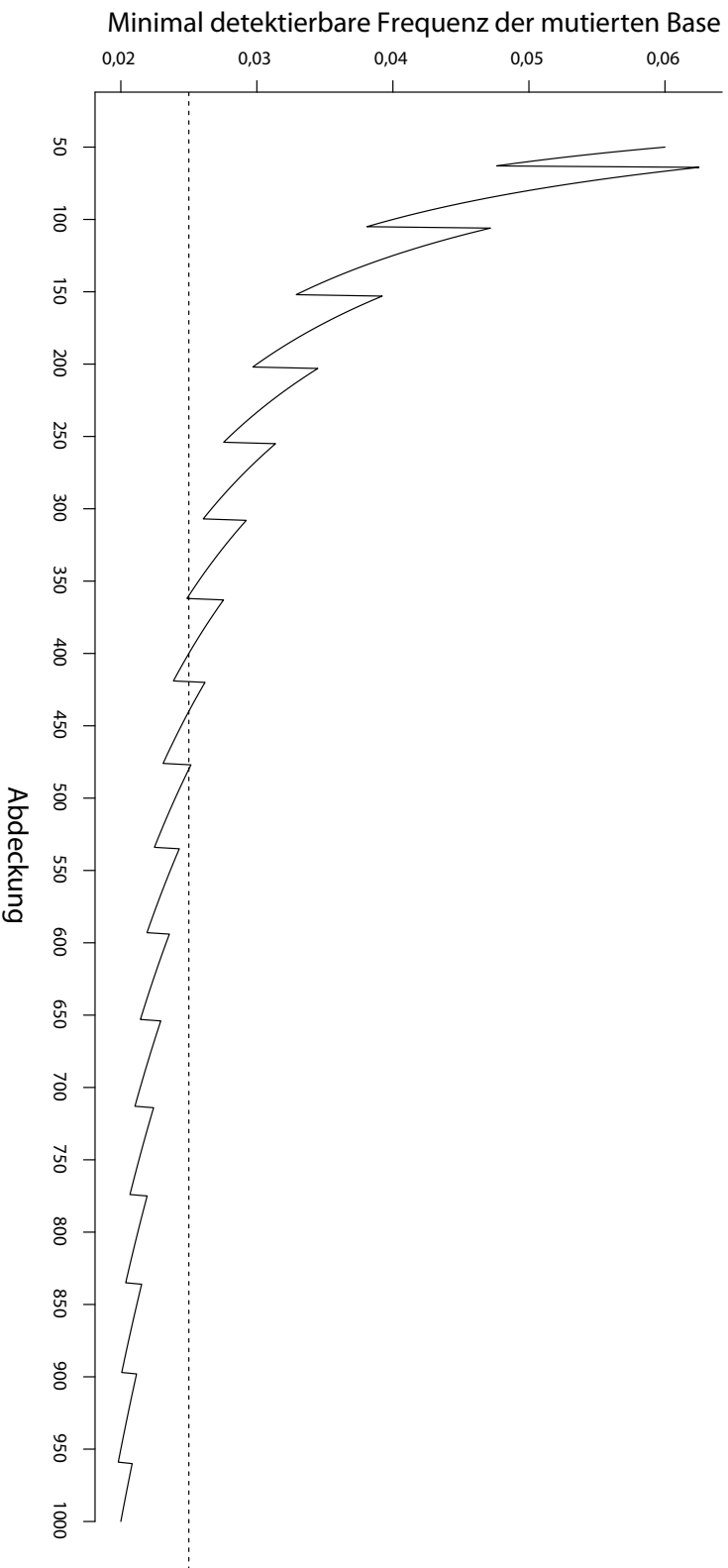


Abbildung 5.9: Darstellung der minimal detektierbaren Frequenz einer mutierten Base in Abhängigkeit der vorliegenden Abdeckung. Es wurde jeweils die minimale Anzahl mutierter Basen bestimmt, bei der der Binomialtest für die vorliegende Abdeckung einen p-Wert $< 5\%$ ergab. Die Treppenstufen sind das Ergebnis der ganzzahligen „Sprünge“ mutierter Basen. Bei einer Abdeckung von 63 NGS-Sequenzfragmenten ist beispielsweise eine Anzahl von 3 mutierten Basen erforderlich (minimal Frequenz $= \frac{3}{63} = 4,76\%$); bei einer Abdeckung von 64 NGS-Sequenzfragmenten sind hingegen 4 mutierte Basen erforderlich, um einen p-Wert $< 5\%$ zu erzielen (minimal Frequenz $= \frac{4}{64} = 6,25\%$).

5.2.4 Binomialansatz für artifizielle, heteroresistente NGS-Proben

Der Binomialansatz wurde mit dem Statistikprogramm R (Kapitel 2.5.1) umgesetzt. Der Aufruf ist im Quelltext 5.2 dargestellt. Bei dem Aufruf beschreibt x_i die Anzahl mutierter Basen an Position i , n_i die Abdeckung an der Position i und p_i die Fehlerwahrscheinlichkeit nach Basenqualitätswert-Transformation (Gleichung 5.1) an der Stelle i . Die Fehlerwahrscheinlichkeit an der Stelle i wurde jeweils, wie in Quelltext 5.3 beschrieben, berechnet.

```
binom.test((x_i, n_i, p_i, alternative="greater", conf.level=0.95)
```

Quelltext 5.2: Aufruf des exakten binomischen Tests mit R Version 3.0.2

```
quality<-c(qualA, qualC, qualG, qualT)
meanerror<-10^(-mean(quality[which(quality>0)])/10)
```

Quelltext 5.3: Berechnung der Fehlerwahrscheinlichkeit anhand des Basenqualitätswerts mit R Version 3.0.2

Die NGS-Sequenzfragmente der FastQ-Dateien aus den artifiziellen, heteroresistenten Mischproben (Tabelle 2.3) wurden zur Referenzsequenz *M. tuberculosis* H37Rv (GenBank NC000962.3) ausgerichtet, Duplikate entfernt, um Indels neu ausgerichtet und die Basenqualitätswerte rekaliert (Analyseschritten eins bis fünf aus Abbildung 2.3). Die entstandenen BAM-Dateien wurden anschließend mit dem erstellten Perl-Skript „Batch_BamBasecount_pvalue.pl“ (Quelltext: https://github.com/vschleusener/Skripte_vschleusener) analysiert. Der allgemeine Ablauf des Skriptes ist in Abbildung 5.10 dargestellt. Zuerst wird der Bam-readcount Algorithmus aufgerufen, die Anzahl nicht übereinstimmender Basen pro Position mit dem Binomialtest bewertet und anschließend in eine Tabelle konvertiert. Die Analyse wurde auf bekannte, resistenzvermittelnde Positionen, die auch für die Resistenzvorhersage der Webseite PhyResSE verwendet werden, begrenzt (Anhang D). Der minimale Basenqualitätswert für die Berechnung des Bam-readcount Algorithmus wurde auf Q20 (< 1 % Fehlerwahrscheinlichkeit) gesetzt (Kapitel 2.2.1, Abbildung 2.2).

In Tabelle 5.8 ist das Ergebnis des Binomialtests für die Proben SR4k_rpoB526_1, SR4k_rpoB526_5 und SR4k_rpoB526_10 dargestellt. Es wurde jeweils nur die gesuchte, resistenzvermittelnde Mutation 761140 A → C des Selektionsklons SR4k mit einem p-Wert unter 0,05 detektiert. Die Frequenz des alternativen Allels

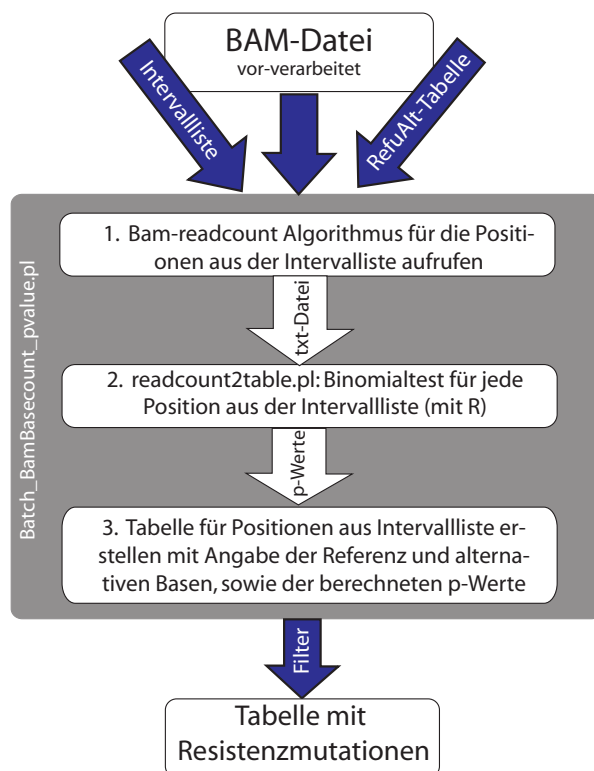


Abbildung 5.10: Schematische Darstellung des Perl-Skripts `Batch_BamBasecount_pvalue.pl` zur Bewertung niederfrequenter Resistenzmutationen (Quelltext: https://github.com/vschleusener/Skripte_vschleusener). Das Skript wird im Ordner der zu untersuchenden BAM-Dateien aufgerufen. Die Übergabeparameter sind eine Intervallliste, die die zu untersuchenden Positionen definiert und eine zugehörige Tabelle (`RefuAlt`) mit Angaben zum Referenz- sowie zum Alternativallel der Positionen aus der Intervallliste. Zuerst wird der Bam-readcount-Algorithmus für die definierten Positionen der Intervallliste auf den vor-verarbeiteten BAM-Dateien aufgerufen. Anschließend wird der Binomialtest für alle Positionen durchgeführt und eine Tabelle erzeugt. Zuletzt werden die Positionen anhand eines vorgegebenen p-Werts (hier 5%) gefiltert. Die Ausgabe ist eine Tabelle mit den detektierten, resistenzvermittelnden Mutationen.

Tabelle 5.8: Detektierte, resistenzvermittelnde Mutationen der artifiziellen NGS-Mischproben SR4k_rpoB526_1, SR4k_rpoB526_5 und SR4k_rpoB526_10 mit Hilfe des Binomialtests. Es konnte jeweils ausschließlich die gesuchte Mutation 761140 A → C des Selektionsklons SR4k detektiert werden. Die Frequenz des alternativen Allels stimmte jeweils mit geringer Abweichung mit der Frequenz des Selektionsklons überein (Tabelle 2.3).

Probe	POS	REF	ALT	DP	Freq(ALT)	Q(ALT)	p-Wert
SR4k_rpoB526_1	761140	A	C	394	0,0076	36,67	$5,99 * 10^{-04}$
SR4k_rpoB526_5	761140	A	C	443	0,0542	35,42	$5,67 * 10^{-35}$
SR4k_rpoB526_10	761140	A	C	586	0,0819	35,58	$1,14 * 10^{-82}$

Abkürzungen: ALT-alternatives Allel, DP-Abdeckung (engl.: depth), Freq(ALT)-Frequenz des alternativen Allels, POS-Position, Q(ALT)-mittlerer Basenqualitätswert der alternativen Allele, REF-Referenzallel.

stimmte mit geringer Abweichung mit der Frequenz des Selektionsklons für die jeweilige NGS-Probe (1 %, 5 % und 10 %) überein und die mittlere Basenqualität der alternativen Basen war mit Werten über Q35 (0,03 % Fehlerwahrscheinlichkeit; Tabelle 2.1) vergleichsweise gut (Tabelle 5.8).

Die Ergebnisse des Binomialtests für die Proben SR1a_rpoB531_1, SR1a_rpoB531_5 und SR1a_rpoB531_10 sind in Tabelle 5.9 dargestellt. Es wurde ebenfalls

Tabelle 5.9: Detektierte, resistenzvermittelnde Mutationen der artifiziellen NGS-Mischproben SR1a_rpoB531_1, SR1a_rpoB531_5 und SR1a_rpoB531_10 mit Hilfe des Binomialtests. Es wurde jeweils ausschließlich die gesuchte Mutation 761155 C → T des Selektionsklons SR1a detektiert. Die Frequenz des alternativen Allels stimmte jeweils ungefähr mit der Frequenz des Selektionsklons überein (Tabelle 2.3).

Probe	POS	REF	ALT	DP	Freq(ALT)	Q(ALT)	p-Wert
SR1a_rpoB531_1	761155	C	T	492	0,0203	32,80	$2,1 * 10^{-06}$
SR1a_rpoB531_5	761155	C	T	656	0,0503	31,79	$4,52 * 10^{-53}$
SR1a_rpoB531_10	761155	C	T	582	0,1031	31,52	$3,27 * 10^{-118}$

Abkürzungen: ALT-alternatives Allel, DP-Abdeckung (engl.: depth), Freq(ALT)-Frequenz des alternativen Allels, POS-Position, Q(ALT)-mittlerer Basenqualitätswert der alternativen Allele, REF-Referenzallel.

jeweils nur die gesuchte, resistenzvermittelnde Mutation 761155 C → T des Selektionsklons SR1a mit einem p-Wert unter 0,05 detektiert. Die Frequenz der alternativen Allele stimmten mit geringer Abweichung mit der enthaltenen Fre-

quenz des Selektionsklons für die jeweilige artifizielle NGS-Probe überein und die mittlere Basenqualität der alternativen Allele war mit einem Wert von circa Q32 (0,06 % Fehlerwahrscheinlichkeit; Tabelle 2.1) vergleichsweise hoch (vgl. Tabelle 5.9).

Mit Hilfe der artifiziell hergestellten, heteroresistenten NGS-Proben konnte der Binomialansatz zur Detektion von niederfrequenten SNPs in Resistenzgenen validiert werden. In den folgenden Abschnitten wurde diese Methode auf NGS-Daten von seriellen Proben von TB-Patienten angewendet.

5.2.5 Auswertung serieller Proben von TB-Patienten

In diesem Abschnitt wurden serielle Proben von TB-Patienten aus einer Studie, die in Usbekistan zwischen 2003 und 2008 durchgeführt wurde, mit Hilfe des Binomialtests auf Heteroresistenzen untersucht [36]. Diese Studie behandelte die Entwicklung von OFX-Resistenz sowie die Entwicklung einer XDR-TB während einer MDR-TB-Behandlung. Mit Hilfe der IS6110 DNS Fingerprintmethode wurden die Proben auf Reinfektion sowie Mischinfektion kontrolliert (Kapitel 1.7). Von den insgesamt 87 eingeschlossenen Patienten entwickelten 18 während der Behandlung eine Resistenz gegenüber OFX. Bei vier dieser Patienten konnte in den späteren Proben ein weiterer Stamm identifiziert werden, einer hatte eine Mischinfektion und 13 Patienten waren über die Zeit laut IS6110 DNS Fingerprintmethode mit demselben Stamm infiziert. Von den seriellen Proben dieser 13 Patienten wurden NGS-Daten generiert. Im folgenden Abschnitt werden die Ergebnisse der NGS-Analyse mit Hilfe des zuvor beschriebenen Binomialtests sowie des Mischungsdetektors beschrieben.

In Tabelle 5.10 sind die phänotypischen DST-Ergebnisse sowie die Anzahl der NGS-Datensätze für diese 13 Patienten aufgetragen. Die phänotypischen DST-Ergebnisse zeigen neben der Entwicklung von OFX-Resistenz stark variierende Resistenzprofile zu den unterschiedlichen Zeitpunkten (Tabelle 5.10).

Die NGS-Daten der seriellen Proben wurden, wie in Abbildung 2.3 beschrieben, prozessiert. Die vor-verarbeiteten BAM-Dateien wurden anschließend an den bekannten, resistenzvermittelnden Positionen der Webseite PhyResSE (Anhang D) mit dem Perl-Skript Batch_BamBasecount_pvalue.pl (Abbildung 5.10) auf niederfrequente SNPs untersucht. Auf diese Weise wurde überprüft, ob in den initialen

Tabelle 5.10: Phänotypische DST-Resistenzprofile im Laufe der Behandlung von 13 Patienten aus Usbekistan, die laut IS6110 DNS Fingerprintmethode mit dem gleichen Stamm infiziert waren und OFX-Resistenz während der Behandlung entwickelten. Modifiziert aus [36].

Pat.	Resistenzprofil initial	Resistenzprofil 1 ¹	M.	Resistenzprofil 2 ²	M.	#NGS- Daten- sätze
A	INH, EMB, RMP, SM, PZA	INH, RMP, SM, PZA, OFX	4	INH, EMB, RMP, SM, PZA,	13	3
B	INH, EMB, RMP, SM, PZA, AMK, CPM	INH, EMB, RMP, SM, PZA, AMK/CPM, OFX	9	INH, EMB, RMP, SM, PZA, AMK, CPM, OFX	14	5
C	INH, RMP, SM, PZA, AMK, CPM, ETH	INH, SM, AMK, CPM, OFX	8	-	-	4
D	INH, EMB, RMP, SM, PZA	INH, EMB, RMP, SM, PZA, OFX	7	INH, EMB, RMP, SM, PZA, OFX	15	3
E	INH, EMB, RMP, SM, PZA	INH, EMB, RMP, SM, PZA, OFX	4	-	-	3
F	INH, EMB, RMP, SM, ETH	INH, EMB, RMP, SM, PZA, OFX	4	INH, EMB, RMP, SM, PZA, OFX, AMK, CPM, ETH	14	4
G	INH, EMB, RMP, SM, PZA, ETH	INH, EMB, RMP, SM, PZA, ETH	7	-	-	3
H	INH, RMP, SM, PZA, AMK, CPM	INH, RMP, SM, PZA, CPM, OFX	9	INH, EMB, RMP, SM, PZA, OFX ETH	12	4
I	INH, RMP, SM, ETH	INH, RMP, SM, PZA, ETH, AMK, OFX	9	INH, EMB, RMP, SM, PZA, ETH, AMK, CPM, OFX	18	5
J	INH, RMP, SM, PZA, AMK, CPM	INH, EMB, RMP, SM, PZA, AMK, CPM, ETH, OFX	6	INH, EMB, RMP, SM, PZA, AMK, CPM, ETH, OFX	13	5
K	INH, EMB, RMP, SM, AMK, CPM	INH, EMB, RMP, SM, PZA, AMK, CPM, ETH, OFX	12	INH, EMB, RMP, SM, PZA, AMK, CPM, OFX	16	5
L	INH, RMP, SM, AMK, CPM	INH, EMB, RMP, SM, PZA, AMK, OFX, ETH	3	INH, EMB, RMP, SM, AMK, CPM, OFX	15	4
M	INH, RMP, SM, ETH	INH, EMB, RMP, SM, PZA, ETH, OFX	4	INH, EMB, RMP, SM, PZA, OFX	8	5

Abkürzungen: AMK-Amikacin, CPM-Capreomycin, EMB-Ethambutol, ETH-Ethionamid, INH-Isomiazid, M.-Monat, OFX-Ofloxacin, Pat.-Patient, PZA-Pyrazinamid, RMP-Rifampicin, SM-Streptomycin.

¹ Resistenzprofil bei dem zum ersten Mal eine Resistenz gegen OFX festgestellt wurde

² Letztes Resistenzprofil

NGS-Proben die verantwortliche, resistenzvermittelnde Mutation bereits in geringer Frequenz detektiert werden konnte. Zusätzlich wurden die NGS-Proben mit Hilfe des Mischungsdetektors auf Mischinfektionen geprüft, die zuvor mit der IS6110 DNS Fingerprintmethode nicht unterschieden werden konnten. Venn-Diagrammen der detektierten SNPs zeigten den Verwandtschaftsgrad der MTBK-Stämme unterschiedlicher NGS-Proben desselben Patienten. In Tabelle 5.11 sind die Ergebnisse der Auswertung zusammen gefasst (Anhang F). Basierend auf der NGS-Analyse wurden die Patienten jeweils, sofern möglich, in eine der drei Kategorien unterschieden:

1. Mischinfektion,
2. Reinfektion und
3. Resistenzentwicklung unter Therapie (Wirtsinterne Diversifikation)

zugeordnet.

Bei 9 der 13 Patienten (77 %) konnte eine Re- beziehungsweise Mischinfektion mit unterschiedlichen MTBK-Stämmen detektiert werden. Bei allen Patienten, bis auf zwei Ausnahmen (Patient B und Patient D) handelte es sich um verschiedene Stämme der gleichen Linie (Beijing, Linie 2.2.1). In den zwei Ausnahmen konnte zusätzlich anhand spezifischer SNPs ein Stamm der Linie 4.8 identifiziert werden (Tabelle 5.11).

Mischinfektion

Bei 6 (Patienten B, G, H, K, L und M) der 13 Patienten konnte während des Therapieverlaufs mindestens eine NGS-Probe mit einem heterogenen Resistenzprofil beobachtet werden, das auf eine Mischinfektion hindeutet (Tabelle 5.11). Ein Beispiel für einen Patienten mit einer nachgewiesenen Mischinfektion ist in Abbildung 5.11 dargestellt. In der vierten NGS-Probe des Patienten B wurden eine Mischinfektion mit Hilfe des Mischungsdetektors detektiert und es wurden spezifische SNPs der phylogenetischen Linien 2.2.1 und 4.8 detektiert (Abbildung 5.11). Außerdem wurden zusätzliche, zuvor nicht detektierte, resistenzvermittelnde SNPs detektiert. Die in den vorherigen NGS-Proben detektierten resistenzvermittelnden SNPs wurde mit einer alternative Allelfrequenz von ungefähr 15 % nachgewiesen. Für die verbleibenden fünf Patienten (Patienten G, H, K, L und M)

Tabelle 5.11: Analyse serieller NGS-Daten von 13 Patienten aus Usbekistan, die basierend auf der IS6110 DNS Fingerprintmethode über den Behandlungszeitraum mit dem gleichen Stamm infiziert waren und OFX-Resistenz unter Therapie entwickelten. Für jeden Patienten ist die phylogenetische Klassifizierung der Stämme sowie die Antibiotika, für die anhand der resistenzvermittelnden SNPs in mindestens einer der NGS-Proben eine Heteroresistenz (resistenzvermittelnder SNP mit einer Frequenz $< 75\%$ alternativen Allel) detektiert wurde, dargestellt. Anhand dieser Ergebnisse erfolgte eine Einteilung in die drei Kategorien Mischinfektion, Reinfektion und Resistenzentwicklung unter Therapie vorgenommen.

Patient	Phylo.-Klassifizierung	Heteroresistenz SNP-Freq $< 75\%$	NGS-Auswertung
A	Beijing (Linie 2.2.1)	PZA, OFX	Re- oder Mischinfektion mit 3 Stämmen
B	Beijing (Linie 2.2.1)/ Linie 4.8	INH, SM, AMK/CPM, OFX	Mischinfektion mit 2 Stämmen
C	Beijing (Linie 2.2.1)	RMP	Wirtsinterne Diversifikation (Reversion)
D	Beijing (Linie 2.2.1)/ Linie 4.8	-	Re- oder Mischinfektion mit 2 Stämmen
E	Beijing (Linie 2.2.1)	OFX	Wirtsinterne Diversifikation
F	Beijing (Linie 2.2.1)	OFX	Wirtsinterne Diversifikation
G	Beijing (Linie 2.2.1)	INH, EMB, PZA, AMK/CPM	Mischinfektion mit 2 Stämmen
H	Beijing (Linie 2.2.1)	EMB, PZA, AMK/CPM, OFX	Mischinfektion mit 2 Stämmen
I	Beijing (Linie 2.2.1)	-	Re- oder Mischinfektion mit 3 Stämmen
J	Beijing (Linie 2.2.1)	-	Wirtsinterne Diversifikation
K	Beijing (Linie 2.2.1)	EMB, PZA, AMK/CPM, OFX	Mischinfektion mit 2 Stämmen
L	Beijing (Linie 2.2.1)	OFX	Mischinfektion mit 2 Stämmen
M	Beijing (Linie 2.2.1)	INH	Mischinfektion mit 2 Stämmen

Abkürzungen: AMK-Amikacin, CPM-Capreomycin, EMB-Ethambutol, Freq-Frequenz, INH-Isoniazid, OFX-Ofloxacin, Phylo.-Phylogenetische, PZA-Pyrazinamid, RMP-Rifampicin, SM-Streptomycin.

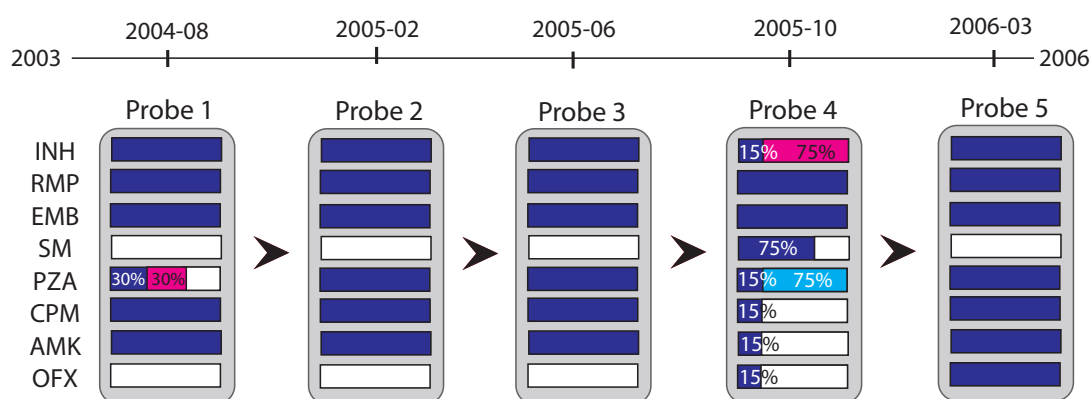


Abbildung 5.11: Entwicklung des Resistenzprofils von Patient B anhand von NGS-Daten. Zwischen Probe 1 und 2 setzt sich eine der beiden in Probe 1 detektierten PZA-resistenten Populationen durch. Die Proben 2 und 3 zeigen dasselbe Resistenzprofil. In der Probe 4 wurden zwei unterschiedliche phylogenetische Linien detektiert und auch das Resistenzprofil zeigt Heteroresistenzen, wobei die zuvor detektierten resistenzvermittelnden SNPs mit 15% alternativem Allel nachgewiesen wurden. Zusätzlich entwickelte sich eine OFX-Resistenz zwischen Probe 3 und Probe 4. Blau = Resistenz-SNP 1, Magenta = Resistenz-SNP 2, Cyan = Resistenz-SNP 3, Weiß = kein SNP

wurde ebenfalls mindestens eine NGS-Probe identifiziert, in der zwei unterschiedliche Stämme gleichzeitig anhand heterogener Resistenzprofile vorkamen (Anhang **F**). Diese NGS-Proben wurden von dem Mischungsdetektor allerdings nicht als Mischungen erkannt.

Reinfektion

Bei 3 (Patienten A, D und I) der 13 Patienten wurden abweichende Resistenzprofile zwischen den seriellen NGS-Proben detektiert. Es konnte gezeigt werden, dass der Patient während des Therapieverlaufs mit unterschiedlichen MTBK-Stämmen infiziert war (Tabelle 5.11).

Bei Patient D wurden die Stämme der unterschiedlichen Linien (Linie 2.2.1 und Linie 4.8) zum Beispiel in separaten NGS-Proben detektiert. Der Stamm der Linie 4.8 wurde ausschließlich in der ersten NGS-Probe beobachtet (Abbildung 5.12). In dem zugehörigen Venn-Diagramm ist zu erkennen, dass die erste Probe von

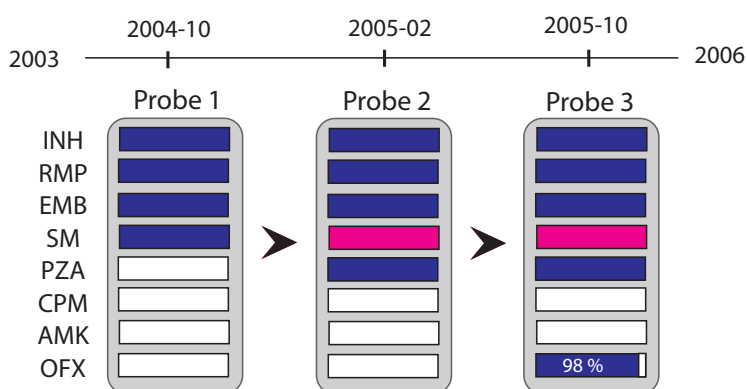


Abbildung 5.12: Entwicklung des Resistenzprofils von Patient D anhand von NGS-Daten. Für die Probe 1 und die Probe 2 wurden unterschiedliche resistenzvermittelnde SNPs für SM sowie ein zusätzlicher resistenzvermittelnder SNP für PZA detektiert. Die Resistenzprofile der Probe 2 und der Probe 3 stimmten überein, bis auf eine OFX-Resistenz, die unter Therapie entwickelt wurde. Blau = Resistenz-SNP 1, Magenta = Resistenz-SNP 2, Weiß = kein SNP

Patient D sich in 801 individuellen SNPs von den anderen beiden Proben unterscheidet (Abbildung 5.13) und eine Reinfektion mit einem Stamm der Linie Beijing (Linie 2.2.1) zwischen den Proben 1 und 2 wahrscheinlich ist.

Die Ergebnisse der NGS-Datenanalyse von Patient A und Patient I weisen darauf hin, dass es sich um eine Infektionen mit drei verschiedenen MTBK-Stämmen derselben phylogenetischen Linie (Beijing, Linie 2.2.1) handelte (Anhang F; Kapitel F.1 und F.9). Für den Patienten I wurden während der NGS-Analyse zum

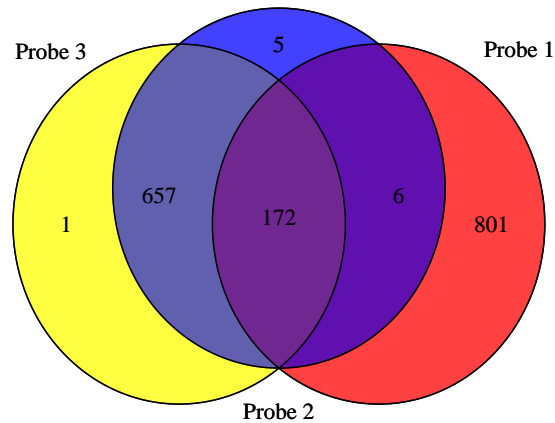


Abbildung 5.13: Venn-Diagramm der SNP-Positionen der VCF-Dateien von Patient D. NGS-Probe 1 (Linie 4.8) unterschied sich mit 801 exklusiven SNPs von den anderen NGS-Proben (Linie 2.2.1).

Beispiel drei verschiedene Resistenzprofile unter Therapie nachgewiesen (Abbildung 5.14). Das Venn-Diagramm der SNPs aus den NGS-Proben von Patient I zeigt ebenfalls, dass sich die Proben 1 und 2 jeweils voneinander und von den übrigen drei Proben unterschieden (Abbildung 5.15). Es bleibt unklar, ob die drei, anhand ihres Resistenzprofils verschiedenen, MTBK-Stämme den Patienten parallel als Mischinfektion oder nacheinander als jeweilige Reinfektion infizierten. Anhand der vorliegenden Ergebnisse ist es nicht möglich für den Patienten D zwischen einer Reinfektion und einer Mischinfektion zu unterscheiden. In den NGS-Proben dieses Patienten wurde jeweils nur ein eindeutiges Resistenzprofil beobachtet, das sich von den anderen NGS-Proben in mindestens einer resistenzvermittelnden Mutation unterscheidet (Anhang F, Abbildung 5.12).

Wirtsinterne Diversifikation

Bei drei Patienten (Patienten C, E und F) wurden keine Hinweise auf unterschiedliche MTBK-Stämme nachgewiesen, sodass für die Resistenzentwicklung nur die wirtsinterne Diversifikation der initialen Population in Frage kommt. Ein Beispiel

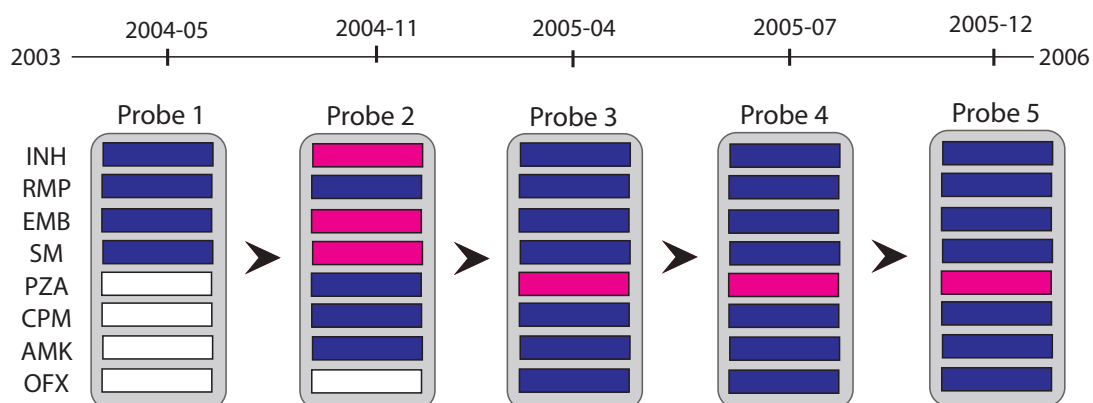


Abbildung 5.14: Entwicklung des Resistenzprofils von Patient I anhand von NGS-Daten. Es konnten drei verschiedene Resistenzprofile im Therapieverlauf detektiert werden. Blau = Resistenz-SNP 1, Magenta = Resistenz-SNP 2, Weiß = kein SNP

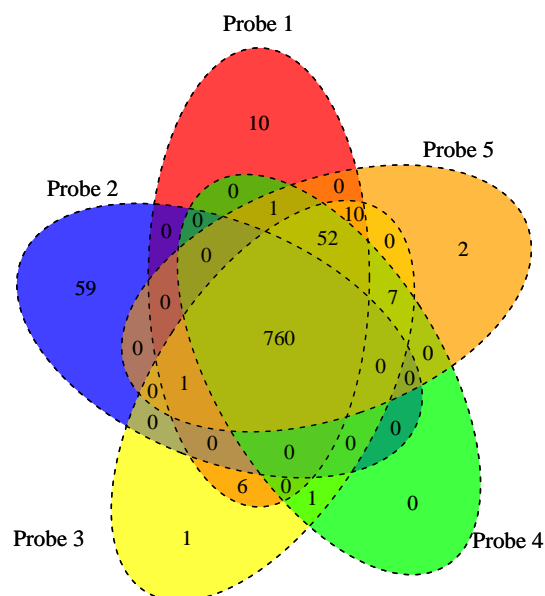


Abbildung 5.15: Venn-Diagramm der SNP-Positionen der VCF-Dateien von Patient I. Für die NGS-Probe 1 wurden 10 SNPs detektiert, die in keiner anderen Probe nachgewiesen wurden. Die NGS-Probe 2 unterscheidet sich mit 59 Mutationen von den anderen NGS-Proben, während für die Proben 3, 4 und 5 nur geringe SNP-Unterschiede detektiert wurden.

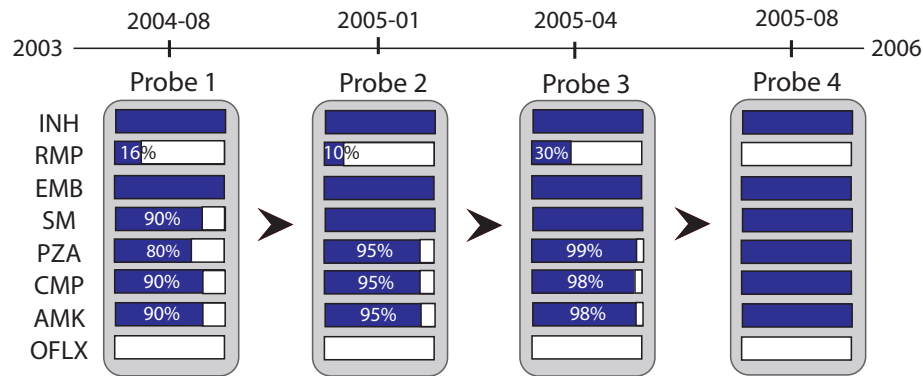


Abbildung 5.16: Entwicklung des Resistenzprofils von Patient C anhand von NGS-Daten. Die Resistenzmutationen für SM-, PZA- und CPM/AMK-Resistenz, die zu Beginn der Behandlung bei 80 % - 90 % der ausgerichteten NGS-Sequenzfragmente detektiert wurden, wurden im Laufe der Behandlung selektioniert und konnten in Probe 4 bei 100 % der ausgerichteten NGS-Sequenzfragmente detektiert werden. Wohingegen die niederfrequente RMP-Resistenzmutation (16 % der ausgerichteten NGS-Sequenzfragmente) der Probe 1 ab Probe 4 nicht mehr detektiert werden konnte. Blau = Resistenz-SNP, Weiß = kein SNP

ist in Abbildung 5.16 dargestellt. Die Resistenzprofile unterscheiden sich zwar zwischen den verschiedenen NGS-Proben, allerdings lassen sich die beobachteten Diskrepanzen auf unterschiedliche Allelfrequenzen derselben resistenzvermittelnden SNPs zurückführen (Abbildung 5.16). Die Abbildung 5.17 zeigt außerdem, dass anhand der Genomdaten zwischen allen NGS-Proben von Patient C nur wenige SNPs Unterschied beobachtet werden konnten.

Zusätzlich konnte bei Patient C eine Reversion der niederfrequenten RMP-Resistenzmutation beobachtet werden. Während der Behandlung mit Zweitrangantibiotika und dem resultierenden Verzicht auf RMP, bildete sich die RMP-Resistenzmutation, die in der ersten Probe bei 16 % der ausgerichteten NGS-Sequenzfragmente detektiert wurde, schrittweise zurück, bis in der letzten Probe weder phänotypische Resistenz noch eine RMP-Resistenzmutation detektiert werden konnte (Abbildung 5.16).

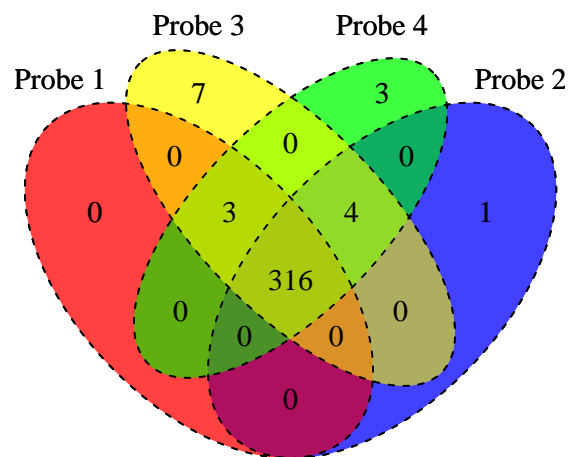


Abbildung 5.17: Venn-Diagramm der SNP-Positionen der VCF-Dateien von Patient C. Die NGS-Proben unterschieden sich in höchstens sieben SNPs voneinander.

Teil III

Diskussion und Ausblick

6 Diskussion

Die TB ist weltweit verbreitet und zählt neben HIV zu den tödlichsten Infektionskrankheiten [213]. Trotz jahrelanger Bemühungen konnte eine Eindämmung der TB nicht erreicht werden, da insbesondere resistente TB die TB-Kontrolle erschweren [159]. Die NGS-basierte Resistenzdiagnostik von MTBK-Stämmen stellt ein vielversprechendes Werkzeug bei der TB-Kontrolle dar, welches allerdings zurzeit unter anderem durch die notwendige, komplexe Analyse der NGS-Daten nicht im klinischen Alltag integriert ist [215]. In dieser Arbeit wurden frei-verfügbare, automatisierte NGS-Analyseprogramme zur NGS-basierten Resistenzdiagnostik von MTBK-Stämmen untersucht. Außerdem wurden spezielle Algorithmen zur verbesserten Analyse von NGS-Daten im Bezug auf heterogene Populationen erstellt. Im Folgenden werden die Ergebnisse dieser Arbeit diskutiert.

6.1 Vergleich automatischer Analyseprogramme für unprozessierte NGS-Daten

In dieser Arbeit wurden die frei-verfügbaren, automatisierten Analyseprogramme CASTB, KvarQ, Mykrobe Predictor TB, TBProfiler, sowie der im Rahmen dieser Arbeit mit entwickelte Webservice, PhyResSE, zur Resistenzbestimmung und Typisierung von MTBK-Stämmen basierend auf unprozessierten NGS-Daten verglichen. Ohne automatisierte Programme ist die Analyse von NGS-Daten auf Grund der notwendigen Kommandozeilen-Algorithmen vorwiegend für Personen mit bioinformatischem Hintergrund zugänglich. Es gibt zwar verschiedene, kommerzielle Software, mit der NGS-Daten analysiert werden können (z. B.: CLC Genomics Workbench, Ridom SeqSphere+ und BioNumerics), allerdings sind diese sehr kostenintensiv und nicht auf MTBK-Stämme und die Resistenzvorhersage spezialisiert [9, 162, 167]. Eine der ersten frei-verfügbaren Webseiten zur automatisierten Resistenzbestimmung und Typisierung von klinischen MTBK-Stämmen

basierend auf NGS-Daten war neben TBProfiler, das parallel entwickelt worden ist, PhyResSE. Zuvor war eine frei-verfügbare NGS-Analyse zur Resistenzvorhersage von MTBK-Stämmen nur von der lokal zu installierenden Software KvarQ möglich, welche bis zum Jahr 2014 allerdings keine Resistenzvorhersage für PZA inkludierte. Allen Programmen gemein ist die Vorhersage der phylogenetischen Linie/Spezies, sowie eine Resistenzvorhersage, basierend auf genetischen Markern. Wesentliche Unterschiede der Programme bestehen neben der verwendeten Technik und dem individuellen Design bei der Ausgabe, der Bedienung und der Datenverarbeitung. Die Klassifizierung der phylogenetischen Linie, aber auch die Resistenzvorhersage sollte für alle Programme von höchster Genauigkeit und auf eine standardisierte Weise erfolgen, was eine Grundvoraussetzung für die schrittweise Etablierung NGS-basierter Diagnostik ist. In dieser Arbeit konnte jedoch gezeigt werden, dass sich die Resistenzvorhersagen der untersuchten Programme für die zwei wichtigsten Erstrangantibiotika (INH und RMP) deutlich unterscheiden, was im für diese Arbeit verwendeten Validierungsdatensatz aus Sierra Leone sowohl zu falsch-positiven als auch zu falsch-negativen Resistenzvorhersagen führte.

PhyResSE ist zurzeit das einzige System, das eine ausführliche Qualitätskontrolle sowohl der unprozessierten NGS-Daten (FastQ-Dateien), als auch der Ausrichtung zur Referenzsequenz (BAM-Dateien) vornimmt. Ein weiteres, bedeutendes Alleinstellungsmerkmal ist die detaillierte Vor-Verarbeitung der NGS-Daten. Neben der Entfernung von duplizierten NGS-Sequenzfragmenten, wird die Ausrichtung der NGS-Sequenzfragmente um kleine Indels herum verbessert und die Basenqualität basierend auf bekannten variablen Stellen rekaliert. Es konnte in dieser Arbeit erstmalig gezeigt werden, dass das Nicht-Durchführen dieser zeitaufwändigen Schritte zwar die Rechenzeit deutlich verringert, allerdings eine große Zahl falsch-positiver Mutationen zur Folge hat. Insgesamt zeigte PhyResSE in diesem Vergleich anhand der resistenz-assoziierten Gene *ahpC*, *embB*, *embA*, *embC*, *gidB*, *inhA*, *katG*, *rpoB*, *rpsL*, *rrs* und *pncA* die beste Sensitivität bei der Resistenzvorhersage mit mindestens 75 % für die verschiedene Antibiotika (PZA). Für die Resistenzvorhersage gegenüber INH zeigte TBProfiler als einziges Programm falsch-positive Resistenzvorhersagen. Ursache dafür ist, dass TBProfiler SNPs im *kasA* Gen in die INH-Resistenzvorhersage einbezieht. Dieses Gen und insbesondere der verwendete SNP (*kasA* G269S) wurde bereits 2007 als nicht direkt mit INH-Resistenz assoziiert beschrieben [197, 198]. Im Jahr 2014, ein

Jahr bevor TBProfiler veröffentlicht wurde, wurde der SNP *kasA* G269S dann als Marker der phylogenetischen Linie LAM (Linie 4.3.3) beschrieben [32, 87]. Als Konsequenz werden alle NGS-Proben, in denen Stämme der Linie 4.3.3 enthalten sind, unabhängig von dem tatsächlichen Resistenzstatus, von TBProfiler als INH-resistent markiert. Diese Tatsache hätte weitreichende Folgen in Regionen, in denen hauptsächlich INH-sensible Stämme der Linie 4.3.3 vorkommen. Bei diesen Patienten müsste auf die Therapie mit Zweitrangantibiotika ausgewichen werden, die nicht nur teurer und zeitintensiver ist, sondern für die Patienten mit starken Nebenwirkungen, wie zum Beispiel die Einschränkung des Farbsehens oder der Hörverlust, verbunden ist. Aus diesem Grund sollte von der Verwendung von TBProfiler zur NGS-basierter Diagnostik (Stand der Kommandozeilen Version 2016) in diesen Regionen der Welt abgesehen werden. In der Webversion wurde dieser Fehler bereits korrigiert, sodass die oben genannte Mutation ausschließlich zur Interpretation der phylogenetischen Linie verwendet wird. Allerdings sind alle Analyseergebnisse der Webseite TBProfiler öffentlich zugänglich, sodass vertrauliche Patientendaten nicht mit der Webversion verarbeitet werden können.

Die diskrepanten Ergebnisse bei der RMP-Resistenzvorhersage der untersuchten Programmen basieren hauptsächlich auf den Unterschieden in den für die Programme hinterlegten Listen resistenzvermittelnder Mutationen. Beispiele sind die validierten Resistenzmutationen *rpoB* T400A und *rpoB* L452P [55, 175], die von CASTB und KvarQ nicht verwendet werden. An dieser Stelle wird deutlich, dass eine standardisierte und vor allem von Experten geprüfte Datenbank mit resistenzvermittelnden Mutationen benötigt wird, um die genombasierte Resistenzvorhersage von MTBK-Stämmen zu vereinheitlichen [111, 112].

Noch größere Unterschiede zwischen den in dieser Arbeit verglichenen Programmen gibt es für die Resistenzvorhersagen von EMB, SM und PZA. Der Grund liegt bei den größeren Genomregionen, in denen Resistenzmutationen auftreten können. Als Folge müssen die Analysealgorithmen größere Genomregionen mit einer größeren Anzahl verschiedener Mutationen abdecken und interpretieren. Für PZA-Resistenz existieren beispielsweise über 400 bekannte, resistenzvermittelnde Mutationen, verteilt über die gesamte Länge des *pncA* Gens (561 bp). Im Gegensatz dazu treten zum Beispiel 95 % der RMP-resistenzvermittelnden Mutationen in der 81 bp großen RRDR des *rpoB*-Gens auf [145]. Bei der PZA-Resistenzvorhersage spielen zusätzlich Indels eine bedeutende Rolle [101, 117], die schwer zu detektieren sind. In dieser Arbeit konnte erstmals gezeigt wer-

den, dass keines der untersuchten Programme eine akkurate Bestimmung und Interpretation von Indels leisten kann. Kleine Indels (< 30 bp) werden zwar von den gängigen Programmen zur Variantendetektion (SAMtools und GATK) detektiert, dennoch können die Start- und Endpositionen nicht akkurat bestimmt werden, sodass eine Interpretation anhand vorgegebener Positionen aus einer Mutationsliste schwierig ist [163]. Noch schwieriger gestaltet sich die Indel-Detektion bei Indels > 30 bp, da diese von den gängigen Ausrichtungsalgorithmen für kurze NGS-Sequenzfragmente nicht akkurat abgebildet werden können [163]. Um zukünftig auch Indels korrekt zu analysieren, sind bereits Algorithmen entwickelt worden [163]. Diese basieren entweder auf einer *de-novo* Assemblierung nicht ausgerichteter NGS-Sequenzfragmente (SOAPindel) [122], der Auftrennung von NGS-Sequenzfragmenten (PRISM und Pindel) [92, 216], der Detektion nicht-abgedeckter Bereiche der Referenzsequenz oder der Analyse der Ausrichtung von „paired-end“ sequenzierten NGS-Sequenzfragmenten mit unerwartetem Abstand zueinander (CNVnator und indelMINER) [2, 163]. Jedes dieser Verfahren hat Vor- und Nachteile [1]. Insbesondere sind sowohl die falsch-positiv, als auch die falsch-negativ Raten bei allen Verfahren ein schwerwiegendes Problem. Eine Optimierung der Algorithmen beziehungsweise deren Weiterentwicklung ist zwingend notwendig. In einem weiterführenden Schritt könnten diese bei den Analyseabfolgen der untersuchten Programme dann integriert werden.

Insgesamt unterschieden sich die Analyseabfolgen der in dieser Arbeit untersuchten Programme deutlich voneinander. Jedoch konnte gezeigt werden, dass sich die Ausgabe für die „sicheren“ SNPs (100 % alternatives Allel) hauptsächlich durch die Interpretation anhand individueller, in den Programmen hinterlegten Mutationslisten unterscheidet. Dieses Problem könnte durch eine einheitliche Datenbank resistenzvermittelnder Mutationen gelöst werden. Doch es konnte auch ein Fall beschrieben werden, bei dem eine Mutation (*embB* G406D [160]) zwar in der Mutationsliste von Mykrobe Predictor TB enthalten war und in der Stampy BAM-Datei eindeutig zu erkennen war, aber von Mykrobe Predictor TB nicht interpretiert und ausgegeben wurde.

Ein weiteres Problem tritt bei der Detektion niederfrequenter SNPs auf. Sobald eine Mutation nicht von der Mehrheit der ausgerichteten NGS-Sequenzfragmente vorhergesagt wird, bedarf es einer Bewertung, ob es sich um eine gehäufte Anzahl an Sequenzierfehlern, ein Ausrichtungsproblem oder tatsächlich um eine heterogene Population handelt. Die verwendeten Programme zur Variantendetektion

(SAMtools und GATK) verwenden ein statistisches Modell und geben nur Mutationen aus, die einen Grenzwert an Sicherheit für die korrekte Detektion überschreiten. Als direkte Folge können Mutationen unterhalb von 10 % alternativem Allel nicht akkurat bestimmt werden [215]. Darüber hinaus unterscheidet sich der Fokus der Analyseabfolgen der untersuchten Programme. Während KvarQ einen schnellen Ansatz mit sicheren Mutationen verfolgt, wurde Mykrobe Predictor TB nach Autorenangaben speziell für die Detektion von niederfrequenten Mutationen implementiert [33].

Bei der Analyse der NGS-Daten des Sierra Leone Datensatzes mit Mykrobe Predictor TB wurden sieben falsch-positive niederfrequente RMP-resistenzvermittelnde SNPs ausgegeben. Diese Positionen überstiegen den Grenzwert von 15 % alternativem Allel, der zur Interpretation nötig ist, zwar nicht, dennoch sollten nur qualitativ hochwertige Positionen ausgegeben werden, um die Fehlinterpretation von Anwenderseite aus zu verhindern. Eine manuelle Analyse der betroffenen Stellen mit Hilfe von Visualisierungsprogrammen zur Darstellung der ausgerichteten NGS-Sequenzfragmente zeigte, dass alle alternativen Basen einen Basenqualitätswert unter 15 besaßen ($> 3\%$ Fehlerwahrscheinlichkeit). Dieses Beispiel verdeutlicht die Notwendigkeit von zusätzlichen Qualitätswerten (Basenqualität, Ausrichtungsqualität, o. ä.) bei der Bewertung von NGS-Daten, insbesondere niederfrequenter Mutationen. Das Problem fehlender Qualitätswerte zeigte sich ebenfalls bei CASTB, KvarQ und TBProfiler. PhyResSE erstellte als einziges Programm einen ausführlichen Qualitätsreport. Jedoch hat auch PhyResSE die artifiziellen, heteroresistenten NGS-Proben nicht detektiert. Die niederfrequenten SNPs wurden nicht detektiert, da der verwendete Algorithmus zur Variantendetektion Mutationen unter 10 % alternativem Allel nicht akkurat nachweist [215]. Für PhyResSE zeigte sich weiterhin, dass durch die strikte Vor-Verarbeitung der BAM-Dateien potentiell richtig-positive Mutationen übersehen werden können. Ebenso wie die Programme CASTB, KvarQ und TBProfiler liegt das Hauptziel von PhyResSE in der Vorhersage von sicheren Resistenzen, nicht in dem Ausschluss von möglichen Resistenzen. Das heißt, das Nicht-Detektieren niederfrequenter Mutationen wird zu Gunsten der Qualität vorhergesagter Mutationen toleriert. Mykrobe Predictor TB klassifizierte in diesem Zusammenhang als einziges der untersuchten Programme die NGS-Probe des Stammes 8867-03 (ENA-Zugriffsnummer: ERS457115) als EMB-resistent anhand einer niederfrequenten Mutation. Auf dem Sanger-Chromatogramm ließ sich ein schwaches Signal beob-

achten, das auf eine niederfrequente Resistenzmutation hindeutete. Dieser Stamm wurde zwar phänotypisch als EMB-sensibel getestet, allerdings zeigt eine aktuelle Studie die Unzuverlässigkeit der phänotypischen Resistenztestung für EMB [128]. Insgesamt ist also für eine valide Resistenzvorhersage die Etablierung, sowie eine ständige Kontrolle und Erweiterung einer validierten SNP-Datenbank von großer Bedeutung. Im Rahmen der ReSeqTB Initiative könnten in Zukunft die Resistenzvorhersagen der untersuchten Programme vereinheitlicht werden [113, 189]. In dieser Arbeit wurde weiterhin die Möglichkeit der Justierbarkeit der hinterlegten Mutationslisten von Benutzerseite aus geprüft. Für CASTB gibt es in der verwendeten Version keine Möglichkeit, die Resistenzvorhersage mit Hilfe neuer Mutationen zu verbessern. Im Gegensatz dazu bieten die anderen untersuchten Programme verschiedene Möglichkeiten. Dabei sind für die Programme KvarQ, Mykrobe Predictor TB und TBProfiler Programmierkenntnisse erforderlich, die bei einer breiten Nutzer-Zielgruppe nicht vorausgesetzt werden können. PhyResSE bietet eine benutzerfreundliche Möglichkeit, die Liste resistenzvermittelnder SNPs für die individuelle Benutzeranalyse zu bearbeiten.

Ein weiteres bedeutendes Ergebnis der vorliegenden Arbeit war der Vergleich der phylogenetischen Klassifizierung der fünf Programme basierend auf spezifischen Markern im Genom. Die Einteilung der MTBK-Stämme des verwendeten Validierungsdatensatz aus Sierra Leone in die sieben Hauptlinien war für alle Programme vergleichbar, obwohl Mykrobe Predictor TB alle sechs im Validierungsdatensatz enthaltenen MTBK-Stämme der Linie 5 (Linie West African I) nicht klassifizieren konnte. Eine weitere Ausnahme konnten für CASTB beobachtet werden. Ein MTBK-Stamm der Linie 4.3.3 (Linie LAM) wurde von CASTB fälschlicherweise als *M. bovis* klassifiziert, was durch die intrinsische PZA-Resistenz von *M. bovis* Stämmen eine Therapie ohne PZA implizieren würde [151] und wäre somit von klinisch-relevanter Bedeutung. Obwohl dieser Stamm von CASTB nicht explizit als PZA-resistent markiert wurde, müsste der Nutzer von einer PZA-Resistenz ausgehen, da keiner der überprüften MTBK-Stämme der Spezies *M. bovis* von CASTB als PZA-resistent markiert wurde. Die intrinsische Mutation *pncA* H57D ist offensichtlich in der hinterlegten Mutationsliste von CASTB nicht abgedeckt. Insgesamt sind alle Programme für Anwender ohne bioinformatisches Vorwissen implementiert worden. Die Art und der Umfang der Ausgabe der Ergebnisse ist für diese Nutzer-Zielgruppe hinsichtlich der Verwendbarkeit und Interpretierbarkeit der Daten von entscheidender Bedeutung. CASTB, Mykrobe Predictor TB

und TBProfiler zeigen eine sehr simplifizierte Ausgabe, indem für die in die Analyse eingeschlossenen Antibiotika tabellarisch der Status (resistent/nicht resistent) ausgegeben wird. Auf diese Weise kann der Anwender besonders einfach erkennen, gegen welche Antibiotika der verwendete MTBK-Stamm resistent ist. Jedoch fehlen bei CASTB neben der Ausgabe des Resistenzstatus, Angaben zu den verantwortlichen resistenzvermittelnden SNPs, was für die Bewertung des Resistenzstatus (starke oder schwache Resistenz/Vorhersagesicherheit), insbesondere für Kliniker essentiell ist. Alle anderen geprüften Programme stellen diese grundlegende Information zur Verfügung.

Die Ausgabe von KvarQ ähnelt einer Kommandozeilenausgabe und könnte somit für die angesprochene Nutzer-Zielgruppe, die graphische Benutzeroberflächen gewohnt ist, unkomfortabel sein. Dennoch sind die Hauptergebnisse (Resistenzstatus und phylogenetische Klassifizierung) übersichtlich dargestellt. Der Resistenzstatus wird als Angabe des Antibiotikums zusammen mit dem detektierten Aminosäureaustausch im entsprechenden Gen ausgegeben.

Mykrobe Predictor TB, PhyResSE und TBProfiler liefern im Gegensatz dazu eine sehr benutzerfreundliche Oberfläche, auf der die Ergebnisse auf einfache und verständliche Weise dargestellt sind. Eine Verbesserung der Benutzerfreundlichkeit wird hier durch übersichtliche Tabellen, sowie die Verwendung von Farben und Symbolen erreicht [166].

Weiterhin wurde in dieser Arbeit zur Validierung der Programme sowie zum Vergleich der Qualität bei der Resistenzbestimmung anhand von NGS-Daten die Sanger-Sequenzierung als anerkannter Goldstandard heran gezogen [88]. In einer aktuellen Studie wurde systematisch die Sanger-Sequenzierung mit NGS-Verfahren verglichen [14]. Es konnte gezeigt werden, dass NGS-Verfahren deutliche Vorteile zur Sanger-Sequenzierung aufweisen. Beck *et al.* zeigten, dass es wahrscheinlicher ist eine richtig-positive Variante, die mit Hilfe der NGS-Analyse nachgewiesen wurde, mittels Sanger-Sequenzierung fälschlicherweise zu widerlegen, als eine von dem NGS-Verfahren falsch-positiv detektierte Variante mittels Sanger-Sequenzierung zu identifizieren [14]. Auch in der vorliegenden Arbeit konnten Vorteile der NGS-Verfahren gegenüber der Sanger-Sequenzierung aufgezeigt werden, indem niederfrequente Resistenzmutationen mittels NGS-Analyse detektiert wurden, die routinemäßig mittels Sanger-Sequenzierung nicht nachgewiesen werden konnten.

Darüber hinaus wurden Sensitivität und Spezifität der Resistenzvorhersage der

geprüften Programme mit der phänotypischen Resistenztestung als Goldstandard berechnet. Für die Programme KvarQ, Mykrobe Predictor TB und TBProfiler wurde dies bereits in groß angelegten Studien beschrieben [18, 33, 192]. Die Sensitivität und Spezifität der Resistenzvorhersage aus diesen Studien lagen insbesondere für EMB und SM deutlich über den Werten der in dieser Arbeit berechneten Werte. Die Abweichungen könnten mit der geringen Stichprobengröße des hier verwendeten Validierungsdatensatzes aus Sierra Leone ($n = 91$) verglichen mit den Datensätzen der vorherigen Studien ($n = 3529$ für Mykrobe Predictor TB und KvarQ, $n = 792$ für TBProfiler) oder mit der Struktur der verwendeten Datensätze (z. B.: Häufigkeitsverteilung von EMB-Resistenz) zusammen hängen. Im Vergleich zu den Sensitivitäten und Spezifitäten, die basierend auf den Sanger-Sequenzierungsergebnissen berechnet wurden, traten nur geringe Unterschiede zu den Sensitivitäten und Spezifitäten auf, die basierend auf der phänotypischen Resistenztestung berechnet wurden. Diese Abweichungen lassen sich zum Beispiel durch inakkurate phänotypische DST-Ergebnisse bei der Testung von EMB und PZA erklären, sodass dort häufiger richtig-positive Stämme als falsch-positive Stämme detektiert wurden. Als direkte Folge nahm die Sensitivität (richtig-positiv Rate) einzelner Programme ab, beziehungsweise stieg an, wenn zuvor im Vergleich zur Sanger-Sequenzierung einige Mutationen nicht detektiert worden sind [102, 128]. Insgesamt ist der Nutzen von molekularbiologischen Verfahren, speziell der NGS-Daten, in der klinischen Anwendung zurzeit begrenzt, da die Bedeutung einiger Mutationen noch nicht vollständig aufgeklärt ist, so gibt es neben phänotypisch sensiblen MTBK-Stämmen, die bekannte, resistenzvermittelnde Mutationen aufweisen auch phänotypisch resistente MTBK-Stämme, die keine Mutation in den bekannten resistenzvermittelnden Genen aufweisen. PhyResSE bietet die Möglichkeit alle detektierten Mutationen ohne Beschränkung auf bekannte Resistenzgene auszugeben. Mit Hilfe dieser Analyseergebnisse können genomweite Assoziationsstudien durchgeführt werden und detektierte Mutationen mit dem Phänotyp korreliert werden könnten. Auf diese Weise ist es möglich neue Kandidaten für resistenzvermittelnde Mutationen zu definieren, deren klinische Relevanz anschließend überprüft werden kann.

6.2 Detektion heterogener Populationen

6.2.1 Mischungsdetektor

In dieser Arbeit wurde ein Mischungsdetektor implementiert, der basierend auf NGS-Daten von MTBK-Stämmen Mischinfektionen nachweist. Viele Jahre lang wurde angenommen, dass eine ausgeheilte beziehungsweise aktive TB-Infektion das Risiko einer weiteren TB-Infektion senkt [29, 207]. Shin *et al.* und Jasmer *et al.* konnten jedoch zeigen, dass eine zurückliegende TB-Behandlung sogar ein Risikofaktor für Mischinfektionen ist [90, 185]. Mischinfektionen und deren Nachweis stellen die Diagnostik und Epidemiologie vor neue Herausforderungen. Zum einen kann die Koexistenz von sensiblen und resistenten MTBK-Stämmen zu inakkuraten Resistenzprofilen führen, die die Behandlung erschweren und zu einem verschlechterten Behandlungsergebnis beitragen [149, 168, 200]. Zum anderen können Übertragungen zwischen Patienten nicht nachgewiesen werden, wenn zum Beispiel nur ein Stamm übertragen wird, der im Indexpatienten nicht identifiziert worden ist [66]. Erste Studien von Eyre *et al.* aus dem Jahr 2013 zum Thema Mischinfektionen verwendete einen Maximum Likelihood Ansatz zur Detektion von Mischinfektionen anhand von Genomdaten bei *Clostridium difficile* (*C. difficile*) [54]. Ihr Ansatz basierte auf der Anzahl hoch qualitativer Basen an Positionen, die für ihre Variabilität zwischen unterschiedlichen Stämmen bekannt sind. Für das statistische Modell wurden Metadaten, zum Beispiel zum Krankenhauskontakt mit einbezogen. Dieser Ansatz zeigte sich jedoch als ungeeignet für die Analyse von NGS-Datensätzen der Isolate einzelner Patienten, von denen eventuell nur eine mangelhafte Dokumentation der Krankheitsgeschichte existiert.

Köser *et al.* zeigten 2013 erstmals für TB, dass eine erhöhte Anzahl heterogener Mutationen auf eine Mischinfektion hindeutet. Anhand von 209 heterogenen Positionen wurde in dieser Studie eine Mischinfektion zweier unabhängiger MTBK-Stämme der Linie Beijing gezeigt [203]. Ebenfalls konnten Chan *et al.* 2013 anhand von 398 heterogenen Mutationen bei einer 215 Jahre alten Mumie eine Mischinfektion nachweisen [26].

Drei Jahre später erstellten Gan *et al.* 2016 anhand von 782 NGS-Datensätzen von MTBK-Stämmen aus öffentlichen Datenbanken einen phylogenetischen Referenzbaum (phylogenetische Datenbank) basierend auf 68 639 SNPs [66]. Mischinfektionen wurden identifiziert, indem anhand der individuellen SNPs einer

Probe passende Pfade im Referenzbaum nachgewiesen wurden. Es wurde von einer Mischinfektion ausgegangen, sobald mehr als ein Pfad im Referenzbaum, durch heterogene Positionen der Probe, identifiziert werden konnte. Dieser Ansatz ermöglicht eine Unterscheidung von sehr nah verwandten MTBK-Stämmen, bei einer minimalen Abdeckung von einem NGS-Sequenzfragment für die Subpopulation. Der Referenzbaum muss jedoch kontinuierlich weiter entwickelt werden, da Mischungen nicht nachgewiesen werden können, wenn die divergenten Pfade der zwei MTBK-Stämme nicht im Referenzbaum enthalten sind.

In dieser Arbeit wurde eine neue Methode zum Nachweis von Mischinfektionen implementiert, die auch die Anzahl heterogener Positionen ausnutzt, jedoch erstmalig das Kerngenom und damit den Großteil der gesamten Genominformation der MTBK-Stämme verwendet. In einem repräsentativen, öffentlichen NGS-Datensatz ($n = 1320$, ENA-Zugriffsnummer: ERP000192) konnte gezeigt werden, dass die Sensitivität des hier implementierten Mischungsdetektors für Subpopulationen mit einem Anteil über 5 % an der Gesamtpopulation bei 100 % liegt. Der Grenzwert von 5 % Populationsanteil zur Mischdetektion basiert auf der gewählten Konstruktion des Mischungsdetektors, der eine Variantendetektion mit GATK inkludiert. Für GATK konnte in einer Studie gezeigt werden, dass Mutationen nur bis zu einer Populationsfrequenz von ungefähr 10 % sensitiv detektiert werden und die Werte für Frequenzen darunter deutlich schlechter werden [215]. Bei der Detektion von Mischinfektionen, die auf heterogenen Positionen basiert, besteht die Schwierigkeit in der Unterscheidung von Mikroevolution und Mischinfektion. Aus diesem Grund müssen die MTBK-Stämme einen gewissen genetischen Abstand (z. B.: > 100 heterogene Positionen) voneinander aufweisen [203]. Für den hier verwendeten DBSCAN-Algorithmus wurde ein Minimum von 55 heterogenen Positionen (*MinPts*), basierend auf dem kleinsten Cluster der Testdaten (10 % Population der NGS-Probe Mix90), angenommen, um eine Mischinfektion valide zu detektieren. Das heißt, die an einer Mischinfektion beteiligten Populationen müssen sich in mindestens 55 Positionen unterscheiden, um differenziert werden zu können. Ein Unterschied der bereits bei MTBK-Stämmen derselben phylogenetischen Linie auftreten kann [195].

In der vorliegenden Arbeit wurde davon ausgegangen, dass bei einer Mischinfektion die Frequenz des alternativen Allels ungefähr der Frequenz der Bakterienpopulation im Patienten entspricht. Eine Schwierigkeit bei dieser Annahme besteht darin, dass eine Sputumprobe nicht zwingend die gesamte Bakterienpopulation

des Patienten abdeckt [47, 169]. In dieser Arbeit konnte gezeigt werden, dass die Frequenz des alternativen Allels anhand serieller NGS-Proben von Patienten mit einer Mischinfektion zu den unterschiedlichen Zeitpunkten der Probenentnahme variieren (z. B.: Patient H), sodass auf diese Weise nur der Anteil in der speziellen NGS-Probe bestimmt werden kann. Seriellen Sputumproben zeigen demnach wahrscheinlich nur eine unvollständige Momentaufnahme des tatsächlichen Verhältnisses von einem Stamm A und einem Stamm B in der Patientenlunge. Um die Sputumproben repräsentativer zu gestalten, könnte eine Mischprobe aus Sputumproben von aufeinanderfolgenden Stunden/Tagen kultiviert werden und anschließend die DNS-Bibliothek aus der DNS mehrerer gewachsener Kulturen erstellt werden. Eine weitere Möglichkeit besteht in einem Metagenom-Ansatz, bei dem das komplette Metagenom der Patientenlunge erstellt wird [127].

Bei der Untersuchung verschiedener Populationen wurden die detektierten resistenzvermittelnden SNPs anhand ihrer Frequenz den einzelnen Populationen zugeordnet. Bei einem Großteil der detektierten Mischinfektionen (70 %) mit resistenzvermittelnden SNPs passte die Frequenz der NGS-Sequenzfragmente, die das alternative Allel zeigten, zu den detektierten Häufungsschwerpunkten. Probleme bei der Zuordnung traten auf, wenn zwei Populationen ähnlicher Frequenz (z. B.: beide in 50 %) vorhanden waren, alle SNPs mit 100 % alternativem Allel detektiert wurden oder die Frequenz der alternativen Allele mit keinem der detektierten Häufungsschwerpunkte übereinstimmte. Dies kann der Fall sein, wenn eine der beiden Populationen unabhängig von der Mischinfektion eine wirtsinterne Diversifikation (Mikroevolution) unterläuft und der Anteil in diesem Fall der Subpopulation einer der detektierten Populationen der Mischinfektion entspricht. Fünf der 40 NGS-Proben, die vom Mischungsdetektor in den unterschiedlichen Studienpopulationen als Mischinfektionen detektierten wurden, konnten weder anhand der linienspezifischen SNPs, noch anhand der resistenzvermittelnden SNPs als Mischungen bestätigt werden. Bei diesen NGS-Proben bleibt fraglich, worauf die detektierten Cluster (Ansammlung niederfrequenter SNPs) des DBSCAN-Algorithmus basieren. Bei diesen NGS-Proben wurden Häufungsschwerpunkte zwischen 14 % und 24 % alternativem Allel detektiert. Auch wenn der Mischungsdetektor bei diesen Fällen keine echten Mischinfektionen detektiert hat, wurden auf diese Weise NGS-Proben identifiziert, bei denen eine speziellere Analyse durchgeführt werden sollte, um die ungewöhnlichen Befunde zu begründen.

Bei der Durchführung der Evaluierung des Mischungsdetektors bestand ein Pro-

blem in der Standardprozedur der Probenverarbeitung bei Studien aus Swasiland, Turkmenistan und Nukus. Proben, bei denen im MIRU-VNTR Typisierungsmuster nicht eindeutige Loci beobachtet wurden, wurden initial nicht für die NGS-Analyse verwendet. Deshalb wurden für diese Arbeit weitere 30 Proben, die zuvor von der NGS-Analyse ausgeschlossen wurden, inkludiert. Der Mischungsdetektor konnte für 13 dieser Proben eine Mischinfektion bestätigen. Für 18 Proben hat der Mischungsdetektor keine Mischung nachgewiesen und die MTBK-Stämme konnten eindeutig einer einzigen phylogenetischen Linie zugeordnet werden. Außerdem wurden in dieser Arbeit vier NGS-Proben aus Swasiland ($n=3$) und Nukus ($n=1$) als Mischungen identifiziert, die ein unauffälliges MIRU-VNTR Typisierungsmuster aufwiesen. In dieser Arbeit konnte demnach gezeigt werden, dass der implementierte Mischungsdetektor Vorteile gegenüber der Detektion von Mischinfektionen mittels MIRU-VNTR Typisierung aufweist, indem sowohl mittels MIRU-VNTR Typisierung falsch-positive Mischinfektionen, als auch falsch-negative Mischinfektionen vom Mischungsdetektor identifiziert werden konnten. Die berechneten Anteile von Mischinfektionen in den unterschiedlichen Regionen der Welt variierten in dieser Arbeit zwischen 1 % in Sierra Leone und ungefähr 10 % in Turkmenistan. In vorherigen Studien konnten Frequenzen zwischen 2,8 % und 19 %, je nach Studienregion, beobachtet werden [74]. Mischinfektionen werden hauptsächlich in Regionen mit hohen TB-Inzidenzraten oder Ballungsorten, wie Krankenhäusern und Gefängnissen erwartet [29]. Die Studie mit dem maximal beobachteten Anteil von 19 % Mischinfektionen wurde beispielsweise in einem Stadtteil von Kapstadt, Südafrika, mit hoher TB-Inzidenz (TB-Inzidenz $> 1\,000$) durchgeführt [209]. In der genannten Studie trugen 57 % der Patienten, die mit einem MTBK-Stamm der Linie Beijing infiziert waren, zusätzlich einen MTBK-Stamm einer von Beijing verschiedenen Linie. Mischungen ohne einen MTBK-Stamm der Linie Beijing oder mit zwei unterschiedlichen MTBK-Stämmen der Linie Beijing wurden allerdings nicht nachgewiesen, da eine PCR-basierte Methode verwendet wurde, die beijing- und nicht-beijingspezifische Banden detektierte. In dieser Arbeit konnte jedoch gezeigt werden, dass auch in Regionen mit vergleichsweise niedriger TB-Inzidenz (z. B.: Usbekistan) Mischinfektionen eine Rolle spielen. Die Region mit der zweithöchsten Mischinfektionsrate (Studienpopulation aus Kenia $\approx 9\%$ Mischinfektionen) hat zwar eine moderate TB-Inzidenz (TB-Inzidenz: 264), die Studienpopulation wurde allerdings in städtischen Krankenhäusern rund um Nairobi rekrutiert. An diese Städte grenzen Marginalsiedlungen,

in denen die Armut hoch und der Lebensraum begrenzt ist; Faktoren, die eine Mischinfektion begünstigen [90, 185]. Auf diese Weise könnte ein Bias durch die Probenauswahl entstanden sein und die Ergebnisse somit nicht repräsentativ für die jeweiligen Regionen sein. Ebenso ist dies der Fall in der Studienregion Nokus, in der hauptsächlich MDR-TB Fälle für die Untersuchung herangezogen wurden. Ein weiterer Fokus dieser Arbeit lag auf den phylogenetischen Linien, die in Mischinfektionen auftreten. Eine Studie von Hanekom *et al.* in Khayelitsha, Südafrika aus dem Jahr 2013 zeigte, dass Stämme der phylogenetischen Linien Beijing und Haarlem signifikant häufiger in Mischinfektionen auftreten, als MTBK-Stämme anderer phylogenetischer Linien [74]. In der vorliegenden Arbeit wurden ebenfalls am häufigsten (33 % der Mischungen) MTBK-Stämme der Linien Haarlem und Beijing (49 % der Mischungen) in Mischinfektionen identifiziert. Allerdings konnten MTBK-Stämme aller sechs phylogenetischen Hauptlinien in Mischinfektionen beobachtet werden. Eine Erklärung für die häufige Präsenz von MTBK-Stämmen der Linien Haarlem und Beijing in Mischinfektionen basiert auf der weltweiten Verbreitung von MTBK-Stämmen dieser Linien im Gegensatz zum Beispiel zu den ancestralen Linie (*M. africanum*) [138, 194]. Zu dieser Hypothese passt, dass MTBK-Stämme der Linie LAM als Generalisten, also als Spezies mit einer großen ökologischen Nischenbreite [194], in einem vergleichbaren Anteil von 30 % der Mischungen beobachtet werden konnten.

Zusammenfassend zeigt der im Rahmen dieser Arbeit entwickelte Mischungsdetektor bedeutende Vorteile gegenüber der Mischungsdetektion mittels zum Beispiel der MIRU-VNTR Typisierung. Neben der verbesserten Detektion können mit Hilfe des implementierten Mischungsdetektors erstmals auch die Anteile der verschiedenen Populationen bestimmt und die Einzelpopulationen anhand spezifischer SNPs phylogenetisch klassifiziert werden.

6.2.2 Binomialtest

Ein Schlüssel für eine akkurate und personalisierte Behandlungsstrategie, sowie die Eindämmung der Entstehung neuer Resistenzen und der Verbreitung von resistenten MTBK-Stämmen, ist die frühzeitige Detektion von heteroresistenten Populationen [220]. In dieser Arbeit wurde ein Ansatz zur Detektion und Bewertung von niederfrequenten SNPs mit Hilfe des Binomialtests implementiert. Es

konnten die zwei für die RMP-Resistenzvorhersage wichtigsten Resistenzmutationen (*rpoB* 526 und 531 [143]) valide bis zu einer Frequenz von 1 % alternativem Allel bei einer mittleren Abdeckung von 600 NGS-Sequenzfragmenten detektiert werden. Erstmals konnte damit ein Ergebnis erzielt werden, das mit Detektionsgrenze von 1 % der klassischen, phänotypischen Resistenztestung vergleichbar ist [91]. Die phänotypische Resistenztestung hat bei kleinen Subpopulationen eine Nachweiszeit von bis zu 42 Tage [3, 23]. Durch die Genomsequenzierung und die anschließende Analyse mit Hilfe des hier vorgestellten Binomialtests kann diese Nachweiszeit signifikant reduziert werden. In aktuellen Studien, in denen MTBK-Stämme direkt aus den Sputumproben der Patienten gewonnen und sequenziert werden konnten [21], konnte innerhalb einer Woche ein personalisiertes Resistenzprofil erstellt werden [21]. Auf diese Weise könnte die Therapie bereits in kürzester Zeit angepasst werden, sodass eine Übertragung resistenter Stämme oder die Entwicklung weiterer Resistenzen schnellstmöglich eingedämmt werden kann. Automatisierte molekulare Tests, wie der GeneXpert liefern Ergebnisse zwar bereits innerhalb weniger Stunden, allerdings hat eine Studie gezeigt, dass der GeneXpert resistente Subpopulationen, die weniger als 90 % der Gesamtpopulation ausmachen, *in vitro* nicht detektieren kann [219].

Das Hauptproblem bei der Detektion von niederfrequenter SNPs ist die Unterscheidung echter SNPs von Sequenzierfehlern und Ausrichtungsartefakten. In der Studie von Wyres *et al.* wurden unter anderem die Programme SAMtools und GATK zur Detektion niederfrequenter Mutationen in einem Datensatz mit sehr hoher Abdeckung (> 1000 NGS-Sequenzfragmente) verglichen [215]. Es konnte gezeigt werden, dass SAMtools die geringste Sensitivität hatte. In dieser Studie konnten mit SAMtools 49 % der SNPs mit einer Allelfrequenz von ungefähr 25 % detektiert werden. Im Gegensatz dazu konnte GATK in dieser Studie 94 % der SNPs mit einer Allelfrequenz von ungefähr 10 % nachweisen. Geringere Allelfrequenzen konnten mit GATK nur noch deutlich schlechter detektiert werden. Aus diesem Grund wurde in dieser Arbeit ein neuer Ansatz zum Nachweis niederfrequenter SNPs erstellt.

Zur Unterscheidung von Fehlern und echten niederfrequenten SNPs in NGS-Daten wurden bereits experimentelle [79, 180] und bioinformatische [57, 68, 75, 105, 120, 204] Ansätze entwickelt. Alle diese Ansätze benötigen typischerweise eine spezielle Probenverarbeitung oder andere Modifikationen zur standardisierten NGS-Verarbeitung [188]. Zusätzlich werden häufig nur bestimmte Ziel-Regionen sehr

tief sequenziert, wodurch die Information des restlichen Genoms verloren geht oder in einem zusätzlichen Sequenzlauf erzeugt werden muss.

Ein anderer Ansatz wurde für zwei Studien zur Untersuchung der intra-Patienten Diversität von Enteroviren der Gruppe A (RNA-Virus), Erreger der Hand-Fuß-Mund-Krankheit, verwendet [142, 214]. Für die Validierung von niederfrequenten Polymorphismen in den NGS-Daten der Enteroviren wurde die Binomialverteilung verwendet. In Anlehnung an diese Studien wurde dieser Ansatz für die vorliegende Arbeit modifiziert und zum ersten Mal für niederfrequente Mutationen im Bakteriengenom verwendet. Es konnte gezeigt werden, dass es sich um einen geeigneten Ansatz handelt.

Ein Ergebnis der vorliegenden Arbeit ist die minimale Abdeckung von 400 NGS-Sequenzfragmenten pro Genomposition, die theoretisch zur sicheren Detektion von 2,5 % alternativem Allel nötig ist. Ein ähnlicher Wert mit 500 NGS-Sequenzfragmenten für eine 2 % Subpopulation wurde in der Studie von Spencer *et al.* experimentell ermittelt [188]. Zur Berechnung der Simulation wurde als Fehlerwahrscheinlichkeit der Mittelwert der Fehlerwahrscheinlichkeiten der Basen aller resistenzvermittelnden Positionen verwendet. In der Regel befinden sich resistenzvermittelnde Mutationen in gut konservierten Bereichen des Genoms, in denen sowohl die Sequenzierqualität als auch die Ausrichtungsqualität sehr gut ist. Aus diesem Grund könnten diese Werte in dieser Arbeit zu Gunsten einer sicheren Vorhersage unterschätzt worden sein. Es konnte sogar gezeigt werden, dass es möglich ist, kleinere Frequenzen alternativer Allele in Regionen mit geringerer Abdeckung valide zu detektieren (1 % alternatives Allel bei einer Abdeckung von 400 NGS-Sequenzfragmenten).

Mit Hilfe des erstellten Ansatzes zur Detektion niederfrequenter SNPs sollten in dieser Arbeit Mutationen in den initialen Proben zu Behandlungsbeginn identifiziert werden, die für den Resistenzerwerb unter Therapie verantwortlich sind. Um diese Fragestellung zu untersuchen, wurden serielle Daten aus einer Studie in Usbekistan mit Hilfe der NGS-Analyse und des Binomialtests untersucht. Die untersuchten Proben zeigten ein identisches IS6110 DNS Fingerprint Typisierungsmuster und entwickelten im Verlauf der Behandlung unter anderem eine OFX-Resistenz. Bei Veröffentlichung der Studie wurde basierend auf den Typisierungsergebnissen davon ausgegangen, dass in diesen 13 Patienten eine Evolution und damit Entwicklung neuer resistenter Subpopulationen innerhalb der Patienten stattgefunden hat [36].

Die Ergebnisse dieser Arbeit zeigen, dass ausschließlich bei den NGS-Proben der Patienten E und J eine schrittweise Evolution zur OFX-Resistenz erfolgte. In Patient E konnte in der letzten NGS-Probe eine OFX-Resistenzmutation bei 46 % der ausgerichteten NGS-Sequenzfragmente identifiziert werden. Weiteres Probenmaterial zu einem späteren Zeitpunkt, die eine Mikroevolution manifestieren würde, ist nicht vorhanden. Beim ersten Auftreten der OFX-Resistenzmutation in Patient J konnte diese bereits bei 96 % der alternativen Allele beobachtet werden. Bei den Patienten A, H, I und K wird die OFX-Resistenz nicht entwickelt, sondern durch die Reinfektion mit einem resistenten MTBK-Stamm eingeführt. Für die übrigen Stämme ist dies entweder nicht eindeutig, da die Mutation beim ersten Auftreten mit 99 % oder 100 % alternativem Allel vorhanden waren oder es konnte in keiner der NGS-Proben ein resistenzvermittelnder SNP detektiert werden. Mit Hilfe des hier vorgestellten Verfahrens konnten allerdings bedeutende Unterschiede zu der Analyse mit den Standardverfahren aufgezeigt werden. Bei lediglich zwei der 13 Patienten bestätigte sich die vorherige Annahme der Mikroevolution im Patienten, während in dieser Arbeit gezeigt werden konnte, dass die Situation der anderen Patienten deutlich komplexer ist, als initial vermutet.

Erstmalig konnte in dieser Arbeit die Reversion einer RMP-Resistenzmutation beobachtet werden (Patient C). Für INH und PAS wurde eine Rückbildung von Resistenzmutationen bei Verzicht des entsprechenden Antibiotikums bereits beschrieben [165, 201]. Auf diese Weise gleichen die MTBK-Stämme normalerweise einen eventuellen Fitnessverlust aus. Allerdings handelt es sich bei der hier detektierten Mutation *rpoB* R445D um eine Mutation mit geringem Fitnessverlust, sodass eine Rückbildung zu Gunsten der Fitness unwahrscheinlich ist [16, 64]. Allerdings war die Mutation in diesem Fall in der ersten NGS-Probe noch nicht bei 100 % der Population detektierbar, sodass Bakterien ohne diese Mutation wahrscheinlich selektiert wurden.

Widersprüchliche Ergebnisse im Vergleich zur phänotypischen Resistenztestung konnten hauptsächlich für EMB beobachtet werden. Die Vorhersagewerte der phänotypischen Resistenztestung für EMB liegt insbesondere bei niedrigen Proportionen resistenter Bakterien, deutlich unter denen von INH und RMP [102]. Dies konnte in der vorliegenden Arbeit ebenfalls anhand der NGS-Daten bestätigt werden. Das phänotypische DST-Ergebnis für EMB stimmte bei weniger als 70 % ($n = 37$) der Proben mit der Resistenzvorhersage anhand der NGS-Daten überein. Der vorgestellte Binomialtest zur Detektion niederfrequenter Resistenzmutatio-

nen in NGS-Daten von MTBK-Stämmen stellt somit ein vielversprechendes Werkzeug zur verbesserten Analyse und Interpretation von NGS-Daten dar. Es konnte sowohl ein Vorteil gegenüber der phänotypischen Resistenztestung, als auch gegenüber klassischer Typisierungsverfahren (IS6110 DNS Fingerprint Typisierung) festgestellt werden.

6.3 Fazit und Ausblick

In dieser Arbeit wurden die Vergleichbarkeit automatisierter NGS-Analyseprogramme bei der Resistenzvorhersage von MTBK-Stämmen, die Möglichkeit der Detektion von Mischinfektionen und der Detektion von Mikroevolution der Pathogene im Wirt anhand von NGS-Daten untersucht. Dazu wurde die in Borstel entwickelte, automatisierte Lösung zur Resistenzvorhersage von MTBK-Stämmen, PhyResSE, mit den frei-verfügbaren Programmen CASTB, KvarQ, Mykrobe Predictor TB und TBProfiler zum Vergleich herangezogen. Insbesondere wurde die Qualität der Resistenzvorhersage und phylogenetischen Klassifizierung untersucht. Zusätzlich wurde die Fähigkeit der untersuchten Programme Heteroresistenzen, verursacht durch Mischinfektionen oder Mikroevolution, zu detektieren, geprüft. Es konnten neuartige Verfahren entwickelt werden, die Lösungen für den Nachweis von Heteroresistenz liefern. Diese wurden anschließend anhand von Testdaten validiert und auf klinischen Daten angewendet.

Auf diese Weise konnte das Potential NGS-basierter Verfahren für die phylogenetische Klassifizierung und Resistenzvorhersage von MTBK-Stämmen herausgestellt werden. Wyres *et al.* deklarierten in einer Studie aus dem Jahr 2014, dass NGS-basierte Verfahren die traditionellen Verfahren in Zukunft ersetzen werden [215]. Limitierende Faktoren sind nach Wyres *et al.* jedoch anwenderfreundliche Analyseprogramme und die hohen Analysekosten im Vergleich zu traditionellen Verfahren [215].

Außerdem zeigten Hanekom *et al.* bereits 2013, dass die vollständige Detektion heterogener Populationen eine entscheidende Rolle für den Therapieerfolg spielt und forderten neue Verfahren zur Detektion heterogener Populationen [74].

Im Rahmen dieser Arbeit wurden die zurzeit zur Verfügung stehenden Analyseprogramme kritisch beurteilt und neue Programme zum Nachweis von Heteroresistenzen etabliert. In der aktuellen Version sind die Verfahren als Perl-Skripte zu verwenden. Zukünftig sollen die Algorithmen zur Detektion von Mischinfektionen und niederfrequenten Mutationen in die Analyseabfolge von PhyResSE zu integrieren und so einem breiteren Anwenderspektrum zur Verfügung stehen. Auf diese Weise werden die hier festgestellten Schwachpunkte bei der Detektion niederfrequenter Resistenzmutationen mit PhyResSE gleichzeitig behoben, da sowohl der Mischungsdetektor, als auch die Detektion niederfrequenter SNPs mit dem Binomialtest sensitive, sowie spezifische Leistungen bei der Detektion von

heterogenen Populationen zeigen.

Ein Nachteil des Mischungsdetektors besteht in der Abhängigkeit von dem Variantendetektionsalgorithmus von GATK. Um die Grenze detektierbarer Subpopulationen weiter herabzusetzen, sollen verschiedene Variantendetektionsmethoden überprüft werden. Auf diese Weise könnte insbesondere die Leistung bei kleinen Subpopulationen (Anteil $< 5\%$) weiter verbessert werden.

Bei der Verwendung des Binomialtest zur Detektion niederfrequenter SNPs konnte in dieser Arbeit kein direkter klinischer Anwendungsfall gezeigt werden, da hauptsächlich Re- und Mischinfektionen für das Therapieversagen der dargestellten Fälle ursächlich waren. Aus diesem Grund sollen weitere artifizielle NGS-Datensätze mit unterschiedlichen Resistenzmutationen und weiteren, bekannten Mischverhältnissen erzeugt und mit dem Binomialtest untersucht werden. Außerdem soll der Binomialtest für die Analyse weiterer NGS-Datensätze serieller Isolate verwendet werden, um heterogene Populationen zu detektieren.

Insgesamt leistet diese Arbeit einen entscheidenden Beitrag zur verbesserten und erleichterten Analyse von Hochdurchsatzdaten zur Resistenzvorhersage von MTBK-Stämmen. Diese Arbeit unterstützt somit die Etablierung genombasierter Verfahren in der klinischen Anwendung.

Teil IV

Anhang

A ASCII-Kodierung

Tabelle A.1: ASCII-Zeichen für die Werte 33 bis 126

ASCII Kode	Zeichen	ASCII Kode	Zeichen	ASCII Kode	Zeichen
33	!	65	A	96	`
34	"	66	B	97	a
35	#	67	C	98	b
36	\$	68	D	99	c
37	%	69	E	100	d
38	&	70	F	101	e
39	'	71	G	102	f
40	(72	H	103	g
41)	73	I	104	h
42	*	74	J	105	i
43	+	75	K	106	j
44	,	76	L	107	k
45	-	77	M	108	l
46	.	78	N	109	m
47	/	79	O	110	n
48	0	80	P	111	o
49	1	81	Q	112	p
50	2	82	R	113	q
51	3	83	S	114	r
52	4	84	T	115	s
53	5	85	U	116	t
54	6	86	V	117	u
55	7	87	W	118	v
56	8	88	X	119	w
57	9	89	Y	120	x
58	:	90	Z	121	y
59	;	91	[122	z
60	<	92	\	123	{
61	=	93]	124	
62	>	94	^	125	}
63	?	95	-	126	
64	@				

B Datensätze

B.1 Datensatz aus Sierra Leone

Sierra Leone ist ein westafrikanischer Staat, der an Liberia und Guinea grenzt. Es zählt laut Index der humanen Entwicklung (engl.: Human Development Index; HDI) mit einem Wert von 0,413 zu den gering entwickelten Ländern [43]. Der HDI berechnet sich aus Einzelfaktoren, die Einfluss auf die menschliche Entwicklung eines Landes haben. Dazu gehören beispielsweise die Lebenserwartung bei der Geburt, das Bildungsniveau, sowie das Pro-Kopf-Einkommen. Aus einer großen Zahl solcher Einzelindikatoren wird eine Rangliste errechnet. Der HDI kann maximal den Wert 1,0 annehmen.

Die Entwicklung der TB-Inzidenzraten ist in Abbildung B.1 dargestellt. Seit 2009 ist ein leicht abfallender Trend zu beobachten. Mit 310 Fällen pro 100000 Einwohner ist die TB-Rate dennoch hoch im Vergleich zu anderen Regionen (Abbildung 1.1). Die Studienpopulation bestehend aus 92 Stämmen aus Sierra Leone wurde bereits in verschiedenen Studien verwendet und publiziert [55, 81]. Die Proben stammen von Patienten, die in den Jahren 2003/2004 für eine Nachbehandlung in der westlichen Region und in der Region Kenema eingeliefert wurden. Für diese Proben existieren Ergebnisse aus der traditionellen Resistenztestung. Zusätzlich wurde eine Sanger-Sequenzierung der Resistenzgene *katG*, *rpoB*, *embB*, *rrs*, *rpsL*, *gidB* und *pncA* durchgeführt. Für einige Stämme wurden zusätzlich die Gene *inhA*, *ahpC*, *embA* und *embC* mit der Sanger-Methode sequenziert. Ebenso wurden die klassischen Methoden zur Typisierung verwendet (Kapitel 1.5). Die Populationsstruktur ist hoch divers mit zwei *M. africanum* Linien (West African 1, n = 6; West African 2, n = 14) und neun verschiedenen *M. tuberculosis* Linien (Beijing, n = 4; Cameroon, n = 4; EAI, n = 4; Ghana, n = 1; Haarlem, n = 12; LAM, n = 15; S-type, n = 3; Sierra Leone 1, n = 7; Sierra Leone 2, n = 9; X-type, n = 2). Elf Stämme konnten mit den klassischen Verfahren nicht typisiert werden.

Sierra Leone - Tuberkuloseinzidenz (pro 100000 Einwohner)



Abbildung B.1: Entwicklung der TB-Inzidenzrate in Sierra Leone. Datenquelle: [38]

Das Resistenzprofil, basierend auf der klassischen Resistenztestung auf Festmedium ist in Tabelle B.1 dargestellt. Die Hälfte der Proben (52,2 %) sind resistent gegen mindestens ein Erst-Rang Antibiotikum. Etwa 12 % der Proben zeigen einen multiresistenten Phänotyp.

Die DNA der Proben wurde mit dem NexteraXT Vorbereitungsset für die NGS-Sequenzierung vorbereitet und auf dem Illumina MiSeq Sequenzierer nach Herstellerangaben sequenziert (251 bp oder 301 bp, gepaart). Die Rohdaten der Studie können unter der ENA-Zugriffsnummer „ERP008667“ erreicht werden.

B.2 Datensatz aus Russland

Die Russische Föderation, wie Russland auch genannt wird, ist der flächenmäßig größte Staat der Erde. Mit 144 Millionen Einwohnern auf etwa 17 Millionen km^2 allerdings sehr dünn besiedelt. Der HDI liegt bei 0,798. Russland zählt damit zu den Ländern mit einer hohen menschlichen Entwicklung. Die Daten wurden von ENA (Zugriffsnummer: ERP000192) herunter geladen. Die verwendeten Daten gehören zu einer Studie des Welcome Trust Sanger Instituts zur Untersuchung

	Anzahl (%)
Total	92 (100)
Komplett sensibel	44 (47,8)
Beliebige Resistenz	48 (52,2)
SM Resistenz	22 (23,9)
INH Resistenz	30 (32,6)
RMP Resistenz	15 (16,3)
EMB Resistenz	14 (15,2)
PZA Resistenz	9 (9,8)
Mono-Resistenz	23 (25)
SM allein	14 (15,2)
INH allein	8 (8,7)
RMP allein	1 (1,1)
EMB allein	-
PZA allein	-
MDR	11 (12)

Tabelle B.1: Resistenzprofil der Studienpopulation aus Sierra Leone (Westafrika).
 SM: Streptomycin; INH: Isoniazid; RMP: Rifampicin; EMB: Ethambutol; PZA: Pyrazinamid; MDR: Multiresistent

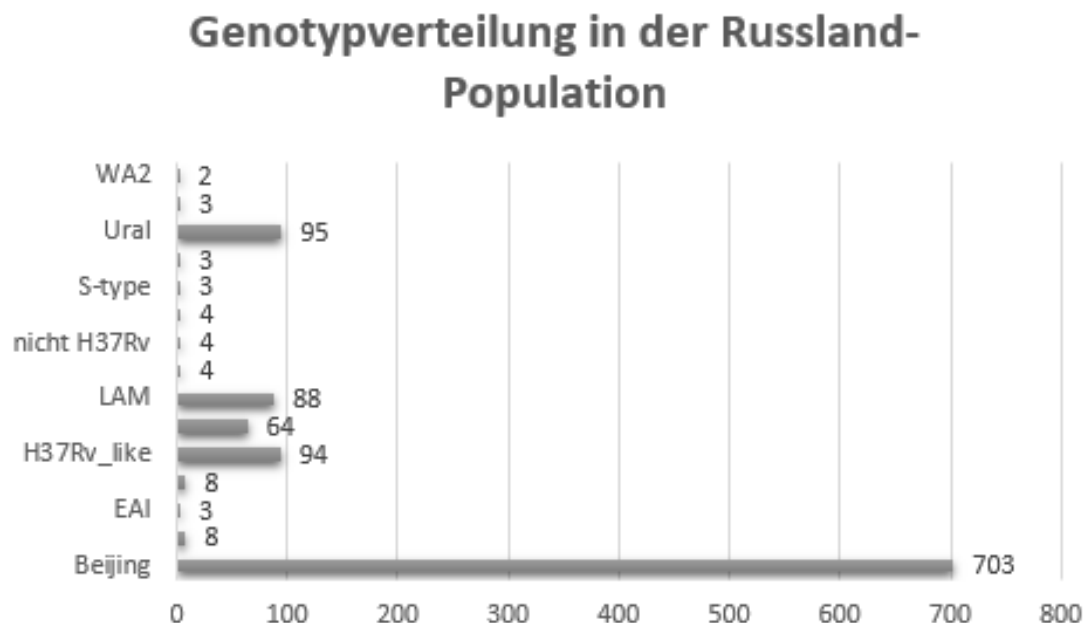


Abbildung B.2: Genotypverteilung der Studienpopulation aus Russland.

der Sequenzdiversität von *M. tuberculosis* in einer russischen Population.

Es wurden prospektiv 1 000 Patientenisolate aus Russland, sowie Referenzdatensätze und XDR-Fälle aus dem Vereinigten Königreich für diese Studie eingeschlossen ($n = 1320$) [24, 25]. Die NGS-Daten wurden mit dem Illumina Genome Analyzer II produziert (gepaart sequenziert). Die Genotypverteilung dieser Population ist in Abbildung B.2 gegeben. Der Großteil der Proben (2/3) gehört der Beijing-Linie an.

Insgesamt zeigte etwa die Hälfte der Proben (48 %) einen multiresistenten Phänotyp. Zusätzlich haben 16 % der MDR-Fälle zusätzliche Resistenzen, sodass diese als XDR definiert sind.

B.3 Datensatz aus Kenia

Die Republik Kenia ist ein Staat in Ostafrika mit ungefähr 45 Millionen Einwohnern und einer Bevölkerungsdichte von 77,56 Einwohnern pro km^2 . Die Studienpopulation besteht aus erwachsenen Patienten mit einer neu-diagnostizierten TB, die in verschiedenen Kliniken in Nairobi rekrutiert wurden. Die Kliniken sind

Rhodes city council Klinik, Aga Khan Krankenhaus, Dagoretti und Waithaka Kliniken.

B.4 Datensatz aus Swasiland

Das Königreich Swasiland befindet sich im südlichen Afrika und hat eine Fläche von nur 17 363 km^2 . Es gehört zu den Ländern mit niedriger menschlicher Entwicklung mit einem HDI von 0,531. Die verwendeten Proben stammen aus einer Querschnittsstudie zur Resistenzüberwachung, die zwischen 2009 und 2010 in Swasiland durchgeführt wurde [174]. Es wurden Patienten aus 15 Studiencenters eingeschlossen. Es wurden sowohl neue Fälle ($n = 352$), als auch Fälle eingeschlossen ($n = 281$), bei denen eine vorherige Behandlung bekannt ist. Bei der Hälfte der vorher behandelten Patienten konnte ein resistenter Phänotyp nachgewiesen werden. Bei den neuen Fällen waren 85 % komplett sensibel (basierend auf DST). Für ungefähr 8 % der neuen Fälle konnte sogar ein MDR-Stamm nachgewiesen werden. Der Anteil von MDR Stämmen lag für behandelte Fälle bei 34 %. Für die NGS-Analyse wurden 280 Proben aus dieser Studie zufällig ausgewählt und mit einem Illumina MiSeq oder einem Illumina NextSeq nach Herstellerangaben sequenziert.

B.5 Datensatz aus Turkmenistan

Turkmenistan ist ein Binnenstaat in Zentralasien und liegt am Kaspischen Meer. Die Bevölkerungsdichte beträgt 13,7 Einwohner pro km^2 . Laut HDI zählt Turkmenistan zu den Ländern mit mittlerer menschlicher Entwicklung (HDI = 0,688). Die Proben stammen aus einer Studie zur Überwachung der von der WHO empfohlenen DOTS-Strategie (Directly Observed Treatment, Short-course) zur Eindämmung der TB-Infektionen. Die DOTS-Strategie wurde im Jahr 1992 eingeführt und dient der konsequenten Betreuung der TB-Patienten. Es wurden 71 MDR-Proben auf einem Illumina MiSeq oder Illumina NextSeq nach Herstellerangaben sequenziert.

B.6 Datensatz aus Usbekistan

Die Republik Usbekistan ist ein islamisch geprägter Binnenstaat in Zentralasien mit einer Fläche von 448 978 km^2 und einer Bevölkerungsdichte von 70 Einwohner pro km^2 . Usbekistan gehört ebenso wie Turkmenistan zu der Regionen mit mittlerer menschlicher Entwicklung (HDI = 0,675). In Usbekistan wurde im Jahr 1998 die DOTS-Strategie eingeführt. Eine Studie zur Resistenzüberwachung in der Region Nukus (Hauptstadt der Autonomen Republik Karakalpakistan in Usbekistan) zwischen 2001 und 2003 wurden hohe Anzahlen MDR-Fälle bei Patienten detektiert, die zuvor Teil des DOTS-Programms waren (40 % der MDR-Fälle). Für diese Arbeit wurden 305 Proben aus dieser Studie verwendet, die auf einem Illumina MiSeq nach Herstellerangaben sequenziert wurden.

C Beispiel für eine Positionsliste des Bam-readcount-Algorithmus

M.tuberculosis_H37Rv	6575	6575
M.tuberculosis_H37Rv	6620	6620
M.tuberculosis_H37Rv	6620	6620
M.tuberculosis_H37Rv	6621	6621
M.tuberculosis_H37Rv	6734	6734
M.tuberculosis_H37Rv	6735	6735
M.tuberculosis_H37Rv	6736	6736
M.tuberculosis_H37Rv	6737	6737
M.tuberculosis_H37Rv	6738	6738
M.tuberculosis_H37Rv	6750	6750
M.tuberculosis_H37Rv	7563	7563
M.tuberculosis_H37Rv	7564	7564
M.tuberculosis_H37Rv	7566	7566
M.tuberculosis_H37Rv	7570	7570
M.tuberculosis_H37Rv	7572	7572
M.tuberculosis_H37Rv	7581	7581
M.tuberculosis_H37Rv	7581	7581
M.tuberculosis_H37Rv	7581	7581
M.tuberculosis_H37Rv	7582	7582
M.tuberculosis_H37Rv	7582	7582
M.tuberculosis_H37Rv	7582	7582
M.tuberculosis_H37Rv	575729	575729
M.tuberculosis_H37Rv	576164	576164
M.tuberculosis_H37Rv	576242	576242
	⋮	

D Mutationslisten

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

Genom-position	Gen	AA Austausch	Anti-bio-tikum	Kvar-Q	Mykr. Pred. TB	Phy-Res-SE	TB-Profler
5508	<i>gyrB</i>	A90G	FQ				
6575	<i>gyrB</i>	R446C	FQ				
6576	<i>gyrB</i>	R446H	FQ				
6576	<i>gyrB</i>	R446L	FQ				
6579	<i>gyrB</i>	S447F	FQ				
6620	<i>gyrB</i>	D461N	FQ				
6620	<i>gyrB</i>	D461H	FQ				
6621	<i>gyrB</i>	D461A	FQ				
6647	<i>gyrB</i>	G470C	FQ				
6648	<i>gyrB</i>	G470A	FQ				
6695	<i>gyrB</i>	I486L	FQ				
6720	<i>gyrB</i>	D494A	FQ				
6734	<i>gyrB</i>	N499Y	FQ				
6734	<i>gyrB</i>	N499D	FQ				
6735	<i>gyrB</i>	N499T	FQ				
6736	<i>gyrB</i>	N499K	FQ				
6737	<i>gyrB</i>	T500P	FQ				
6738	<i>gyrB</i>	T500N	FQ				
6738	<i>gyrB</i>	T500I	FQ				
6741	<i>gyrB</i>	E501V	FQ				

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

6742	<i>gyrB</i>	E501D	FQ				
6749	<i>gyrB</i>	A504T	FQ				
6750	<i>gyrB</i>	A504V	FQ				
6759	<i>gyrB</i>	T507M	FQ				
6767	<i>gyrB</i>	AAC510- GAC	FQ	■			
6768	<i>gyrB</i>	AAC510- GAC	FQ	■			
6853	<i>gyrB</i>	Q538H	FQ				
7247	<i>gyrB</i>	V670F	FQ				
7503	<i>gyrA</i>	R68G	FQ				
7507	<i>gyrA</i>	S69T	FQ				
7510	<i>gyrA</i>	H70R	FQ				
7521	<i>gyrA</i>	A74S	FQ	■			
7554	<i>gyrA</i>	H85X	FQ	■	■		
7555	<i>gyrA</i>	H85X	FQ	■	■		
7556	<i>gyrA</i>	H85X	FQ	■	■		
7557	<i>gyrA</i>	P86X	FQ	■	■		
7558	<i>gyrA</i>	P86X	FQ	■	■		
7559	<i>gyrA</i>	P86X	FQ	■	■		
7560	<i>gyrA</i>	H87X	FQ	■	■		
7561	<i>gyrA</i>	H87X	FQ	■	■		
7562	<i>gyrA</i>	H87X	FQ	■	■		
7563	<i>gyrA</i>	G88X	FQ	■	■		
7563	<i>gyrA</i>	G88C	FQ	■	■	■	
7564	<i>gyrA</i>	G88X	FQ	■	■	■	
7564	<i>gyrA</i>	G88A	FQ	■	■	■	
7565	<i>gyrA</i>	G88X	FQ	■	■	■	
7566	<i>gyrA</i>	D89X	FQ	■	■	■	
7566	<i>gyrA</i>	D89N	FQ	■	■	■	
7567	<i>gyrA</i>	D89X	FQ	■	■	■	
7568	<i>gyrA</i>	D89X	FQ	■	■	■	
7569	<i>gyrA</i>	A90X	FQ	■	■	■	
7570	<i>gyrA</i>	A90X	FQ	■	■	■	
7570	<i>gyrA</i>	A90E	FQ	■	■	■	■

D Mutationslisten

7570	<i>gyrA</i>	A90V	FQ	
7570	<i>gyrA</i>	A90G	FQ	
7570	<i>gyrA</i>	A90V	FQ	
7571	<i>gyrA</i>	A90X	FQ	
7572	<i>gyrA</i>	S91X	FQ	
7572	<i>gyrA</i>	S91P	FQ	
7572	<i>gyrA</i>	S91A	FQ	
7572	<i>gyrA</i>	S91P	FQ	
7573	<i>gyrA</i>	S91X	FQ	
7574	<i>gyrA</i>	S91X	FQ	
7575	<i>gyrA</i>	I92X	FQ	
7576	<i>gyrA</i>	I92X	FQ	
7577	<i>gyrA</i>	I92X	FQ	
7577	<i>gyrA</i>	I92M	FQ	
7578	<i>gyrA</i>	Y93X	FQ	
7579	<i>gyrA</i>	Y93X	FQ	
7580	<i>gyrA</i>	Y93X	FQ	
7581	<i>gyrA</i>	D94X	FQ	
7581	<i>gyrA</i>	D94N	FQ	
7581	<i>gyrA</i>	D94Y	FQ	
7581	<i>gyrA</i>	D94H	FQ	
7581	<i>gyrA</i>	D94H	FQ	
7581	<i>gyrA</i>	D94N	FQ	
7582	<i>gyrA</i>	D94X	FQ	
7582	<i>gyrA</i>	D94V	FQ	
7582	<i>gyrA</i>	D94A	FQ	
7582	<i>gyrA</i>	D94G	FQ	
7582	<i>gyrA</i>	D94G	FQ	
7582	<i>gyrA</i>	D94A	FQ	
7582	<i>gyrA</i>	D94V	FQ	
7583	<i>gyrA</i>	D94X	FQ	
7606	<i>gyrA</i>	P102H	FQ	
7626	<i>gyrA</i>	L109V	FQ	
7677	<i>gyrA</i>	A126R	FQ	
7678	<i>gyrA</i>	A126R	FQ	

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

575729	<i>mshA</i>	Q128Stop	ETH			
576164	<i>mshA</i>	R273C	ETH			
576242	<i>mshA</i>	G299C	ETH			
576338	<i>mshA</i>	Q331Stop	ETH			
576414	<i>mshA</i>	G356D	ETH			
576429	<i>mshA</i>	E361A	ETH			
760314	<i>rpoB</i>	V170F	RMP			
760663	<i>rpoB</i>	A286V	RMP			
761004	<i>rpoB</i>	T400A	RMP			
761033	<i>rpoB</i>	Q409H	RMP			
761074	<i>rpoB</i>	E423A	RMP			
761076	<i>rpoB</i>	F424L	RMP			
761076	<i>rpoB</i>	F424V	RMP			
761077	<i>rpoB</i>	F424S	RMP			
761078	<i>rpoB</i>	F424L	RMP			
761079	<i>rpoB</i>	F425X	RMP			
761080	<i>rpoB</i>	F425X	RMP			
761080	<i>rpoB</i>	TCGGCA- C1274T	RMP			
761081	<i>rpoB</i>	F425X	RMP			
761081	<i>rpoB</i>	CGGCAC- C1275C	RMP			
761082	<i>rpoB</i>	G426X	RMP			
761082	<i>rpoB</i>	G426S	RMP			
761083	<i>rpoB</i>	G426X	RMP			
761083	<i>rpoB</i>	G426D	RMP			
761084	<i>rpoB</i>	G426X	RMP			
761085	<i>rpoB</i>	T427X	RMP			
761085	<i>rpoB</i>	T427S	RMP			
761085	<i>rpoB</i>	T427P	RMP			
761085	<i>rpoB</i>	T427A	RMP			
761086	<i>rpoB</i>	T427X	RMP			
761086	<i>rpoB</i>	T427S	RMP			
761087	<i>rpoB</i>	T427X	RMP			
761088	<i>rpoB</i>	S428X	RMP			

D Mutationslisten

761088	<i>rpoB</i>	S428R	RMP		
761089	<i>rpoB</i>	S428X	RMP		
761089	<i>rpoB</i>	S428I	RMP		
761089	<i>rpoB</i>	S428T	RMP		
761090	<i>rpoB</i>	S428X	RMP		
761090	<i>rpoB</i>	S428R	RMP		
761091	<i>rpoB</i>	Q429X	RMP		
761091	<i>rpoB</i>	Q429K	RMP		
761092	<i>rpoB</i>	Q429X	RMP		
761093	<i>rpoB</i>	Q429X	RMP		
761093	<i>rpoB</i>	Q429H	RMP		
761093	<i>rpoB</i>	Q429H	RMP		
761094	<i>rpoB</i>	L430X	RMP		
761094	<i>rpoB</i>	L430V	RMP		
761095	<i>rpoB</i>	L430P	RMP		
761095	<i>rpoB</i>	L430R	RMP		
761095	<i>rpoB</i>	L430P	RMP		
761095	<i>rpoB</i>	L430R	RMP		
761096	<i>rpoB</i>	L430X	RMP		
761097	<i>rpoB</i>	S431X	RMP		
761097	<i>rpoB</i>	S431R	RMP		
761097	<i>rpoB</i>	S431G	RMP		
761098	<i>rpoB</i>	S431I	RMP		
761098	<i>rpoB</i>	S431T	RMP		
761098	<i>rpoB</i>	S431I	RMP		
761098	<i>rpoB</i>	S431T	RMP		
761099	<i>rpoB</i>	S431X	RMP		
761099	<i>rpoB</i>	S431R	RMP		
761099	<i>rpoB</i>	C1293C- TTCATG	RMP		
761099	<i>rpoB</i>	C1293C- TTC	RMP		
761099	<i>rpoB</i>	CCAATT- C1293C	RMP		
761100	<i>rpoB</i>	Q432K	RMP		

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

761100	<i>rpoB</i>	Q432Stop	RMP	■	■	■
761100	<i>rpoB</i>	Q432E	RMP	■	■	■
761100	<i>rpoB</i>	CAATTC- ATGG- 1294C	RMP	□	■	■
761100	<i>rpoB</i>	Q432K	RMP	■	□	■
761101	<i>rpoB</i>	Q432X	RMP	□	■	□
761101	<i>rpoB</i>	Q432L	RMP	■	■	■
761101	<i>rpoB</i>	Q432P	RMP	■	■	■
761101	<i>rpoB</i>	Q432R	RMP	■	■	□
761101	<i>rpoB</i>	Q432P	RMP	■	■	■
761101	<i>rpoB</i>	Q432L	RMP	■	■	■
761102	<i>rpoB</i>	Q432X	RMP	□	■	□
761102	<i>rpoB</i>	Q432H	RMP	■	■	■
761102	<i>rpoB</i>	A1296A- CGG	RMP	□	□	■
761102	<i>rpoB</i>	ATTCAT- GGAC- 1296A	RMP	□	□	■
761102	<i>rpoB</i>	ATTC- 1296A	RMP	□	□	■
761103	<i>rpoB</i>	F433X	RMP	□	■	□
761103	<i>rpoB</i>	F433L	RMP	■	■	■
761103	<i>rpoB</i>	F433V	RMP	■	■	■
761104	<i>rpoB</i>	F433X	RMP	□	■	□
761104	<i>rpoB</i>	TCAT- 1298T	RMP	□	□	■
761104	<i>rpoB</i>	TCATGG- ACCAGA- A1298T	RMP	□	□	■
761105	<i>rpoB</i>	F433X	RMP	□	■	□
761105	<i>rpoB</i>	F433L	RMP	■	■	■
761105	<i>rpoB</i>	C1299C- TTC	RMP	□	□	■

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

761115	<i>rpoB</i>	N437H	RMP	■	■	■
761115	<i>rpoB</i>	N437D	RMP	■	■	■
761116	<i>rpoB</i>	N437X	RMP	■	■	■
761116	<i>rpoB</i>	N437I	RMP	■	■	■
761116	<i>rpoB</i>	N437T	RMP	■	■	■
761117	<i>rpoB</i>	CAAC- 1311C	RMP	■	■	■
761118	<i>rpoB</i>	N438X	RMP	■	■	■
761119	<i>rpoB</i>	N438X	RMP	■	■	■
761120	<i>rpoB</i>	P439X	RMP	■	■	■
761120	<i>rpoB</i>	N438K	RMP	■	■	■
761120	<i>rpoB</i>	N438K	RMP	■	■	■
761121	<i>rpoB</i>	P439X	RMP	■	■	■
761122	<i>rpoB</i>	P439X	RMP	■	■	■
761123	<i>rpoB</i>	L440X	RMP	■	■	■
761124	<i>rpoB</i>	L440X	RMP	■	■	■
761124	<i>rpoB</i>	L440M	RMP	■	■	■
761124	<i>rpoB</i>	L440L	RMP	■	■	■
761125	<i>rpoB</i>	L440X	RMP	■	■	■
761125	<i>rpoB</i>	L440P	RMP	■	■	■
761126	<i>rpoB</i>	S441X	RMP	■	■	■
761126	<i>rpoB</i>	L440L	RMP	■	■	■
761127	<i>rpoB</i>	S441X	RMP	■	■	■
761128	<i>rpoB</i>	S441L	RMP	■	■	■
761128	<i>rpoB</i>	S441W	RMP	■	■	■
761129	<i>rpoB</i>	G442X	RMP	■	■	■
761130	<i>rpoB</i>	G442X	RMP	■	■	■
761130	<i>rpoB</i>	G442W	RMP	■	■	■
761131	<i>rpoB</i>	G442X	RMP	■	■	■
761131	<i>rpoB</i>	G442A	RMP	■	■	■
761132	<i>rpoB</i>	L443X	RMP	■	■	■
761133	<i>rpoB</i>	L443X	RMP	■	■	■
761134	<i>rpoB</i>	L443X	RMP	■	■	■
761134	<i>rpoB</i>	L443S	RMP	■	■	■
761134	<i>rpoB</i>	L443W	RMP	■	■	■

D Mutationslisten

761135	<i>rpoB</i>	T444X	RMP			
761136	<i>rpoB</i>	T444X	RMP			
761136	<i>rpoB</i>	T444P	RMP			
761137	<i>rpoB</i>	T444X	RMP			
761137	<i>rpoB</i>	T444I	RMP			
761138	<i>rpoB</i>	H445X	RMP			
761139	<i>rpoB</i>	H445N	RMP			
761139	<i>rpoB</i>	H445Y	RMP			
761139	<i>rpoB</i>	H445D	RMP			
761140	<i>rpoB</i>	H445X	RMP			
761140	<i>rpoB</i>	H445L	RMP			
761140	<i>rpoB</i>	H445P	RMP			
761140	<i>rpoB</i>	H445R	RMP			
761140	<i>rpoB</i>	H445P	RMP			
761140	<i>rpoB</i>	H445R	RMP			
761140	<i>rpoB</i>	H445L	RMP			
761141	<i>rpoB</i>	K446X	RMP			
761141	<i>rpoB</i>	CAAG- 1335C	RMP			
761141	<i>rpoB</i>	H445Q	RMP			
761142	<i>rpoB</i>	K446X	RMP			
761142	<i>rpoB</i>	K446Q	RMP			
761143	<i>rpoB</i>	K446X	RMP			
761144	<i>rpoB</i>	R447X	RMP			
761144	<i>rpoB</i>	K446N	RMP			
761145	<i>rpoB</i>	R447X	RMP			
761146	<i>rpoB</i>	R447X	RMP			
761146	<i>rpoB</i>	R447H	RMP			
761146	<i>rpoB</i>	R447P	RMP			
761147	<i>rpoB</i>	R448X	RMP			
761148	<i>rpoB</i>	R448X	RMP			
761149	<i>rpoB</i>	R448X	RMP			
761154	<i>rpoB</i>	S450A	RMP			
761155	<i>rpoB</i>	S450X	RMP			
761155	<i>rpoB</i>	S450L	RMP			

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

761155	<i>rpoB</i>	S450W	RMP	■	■	■
761156	<i>rpoB</i>	S450X	RMP	■	■	■
761157	<i>rpoB</i>	A451X	RMP	■	■	■
761158	<i>rpoB</i>	A451X	RMP	■	■	■
761159	<i>rpoB</i>	A451X	RMP	■	■	■
761160	<i>rpoB</i>	L452X	RMP	■	■	■
761161	<i>rpoB</i>	L452P	RMP	■	■	■
761162	<i>rpoB</i>	L452X	RMP	■	■	■
761166	<i>rpoB</i>	P454S	RMP	■	■	■
761167	<i>rpoB</i>	P454H	RMP	■	■	■
761167	<i>rpoB</i>	P454L	RMP	■	■	■
761167	<i>rpoB</i>	P454R	RMP	■	■	■
761176	<i>rpoB</i>	L457P	RMP	■	■	■
761185	<i>rpoB</i>	E460G	RMP	■	■	■
761196	<i>rpoB</i>	L464M	RMP	■	■	■
761220	<i>rpoB</i>	S472A	RMP	■	■	■
761244	<i>rpoB</i>	I480V	RMP	■	■	■
761245	<i>rpoB</i>	I480T	RMP	■	■	■
761248	<i>rpoB</i>	E481G	RMP	■	■	■
761250	<i>rpoB</i>	T482P	RMP	■	■	■
761254	<i>rpoB</i>	P483L	RMP	■	■	■
761274	<i>rpoB</i>	L490V	RMP	■	■	■
761277	<i>rpoB</i>	I491F	RMP	■	■	■
761284	<i>rpoB</i>	S493L	RMP	■	■	■
761460	<i>rpoB</i>	R552C	RMP	■	■	■
761582	<i>rpoB</i>	E592D	RMP	■	■	■
762089	<i>rpoB</i>	E761D	RMP	■	■	■
764363	<i>rpoC</i>	G332R	RMP	■	■	■
764669	<i>rpoC</i>	P434A	RMP	■	■	■
764670	<i>rpoC</i>	P434R	RMP	■	■	■
764724	<i>rpoC</i>	F452S	RMP	■	■	■
764817	<i>rpoC</i>	V483A	RMP	■	■	■
764817	<i>rpoC</i>	V483G	RMP	■	■	■
764819	<i>rpoC</i>	W484G	RMP	■	■	■
764822	<i>rpoC</i>	D485H	RMP	■	■	■

D Mutationslisten

764822	<i>rpoC</i>	D485N	RMP	■		■
764840	<i>rpoC</i>	I491V	RMP			■
764841	<i>rpoC</i>	I491T	RMP			■
764918	<i>rpoC</i>	V517L	RMP	■		■
764948	<i>rpoC</i>	L527V	RMP	■		■
765462	<i>rpoC</i>	N698S	RMP	■		■
765463	<i>rpoC</i>	N698K	RMP	■		■
765572	<i>rpoC</i>	D735N	RMP			■
765609	<i>rpoC</i>	D747A	RMP			■
766022	<i>rpoC</i>	I885V	RMP			■
781585	<i>rpsL</i>	R9H	SM			■
781678	<i>rpsL</i>	T40I	SM			■
781680	<i>rpsL</i>	T41S	SM			■
781681	<i>rpsL</i>	T41S	SM			■
781687	<i>rpsL</i>	K43T	SM			■
781687	<i>rpsL</i>	K43R	SM	■	■	■
781712	<i>rpsL</i>	K51N	SM			■
781714	<i>rpsL</i>	V52G	SM			■
781810	<i>rpsL</i>	G84V	SM			■
781815	<i>rpsL</i>	R86W	SM			■
781816	<i>rpsL</i>	R86P	SM			■
781818	<i>rpsL</i>	V87L	SM			■
781821	<i>rpsL</i>	K88Q	SM			■
781822	<i>rpsL</i>	K88M	SM	■		■
781822	<i>rpsL</i>	K88T	SM			■
781822	<i>rpsL</i>	K88R	SM	■	■	■
781836	<i>rpsL</i>	V93M	SM			■
801268	<i>rplC</i>	C154R	LZD			■
1416212	<i>embR</i>	Q379R	EMB			■
1416242	<i>embR</i>	I369T	EMB			■
1416282	<i>embR</i>	R356G	EMB			■
1416299	<i>embR</i>	V350G	EMB			■
1416371	<i>embR</i>	S326N	EMB			■
1416390	<i>embR</i>	L320V	EMB			■
1416474	<i>embR</i>	L292M	EMB			■

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

1416630	<i>embR</i>	G240R	EMB						
1416660	<i>embR</i>	R230W	EMB						
1416679	<i>embR</i>	D223E	EMB						
1416717	<i>embR</i>	L211M	EMB						
1416749	<i>embR</i>	E200A	EMB						
1416784	<i>embR</i>	S188R	EMB						
1416786	<i>embR</i>	S188R	EMB						
1416788	<i>embR</i>	A187V	EMB						
1416792	<i>embR</i>	R186G	EMB						
1416821	<i>embR</i>	A176V	EMB						
1416833	<i>embR</i>	L172R	EMB						
1417191	<i>embR</i>	D53H	EMB						
1417254	<i>embR</i>	T32P	EMB						
1417328	<i>embR</i>	V7G	EMB						
1472035	<i>rrs</i>	G190A	SM						
1472122	<i>rrs</i>	G277C	SM						
1472272	<i>rrs</i>	G427T	SM						
1472272	<i>rrs</i>	G427C	SM						
1472307	<i>rrs</i>	C462T	SM						
1472335	<i>rrs</i>	C492A	SM						
1472336	<i>rrs</i>	C492G	SM						
1472337	<i>rrs</i>	C491A	SM	■					
1472337	<i>rrs</i>	C491G	SM	■					
1472337	<i>rrs</i>	C491T	SM	■					
1472337	<i>rrs</i>	C492T	SM	■					
1472358	<i>rrs</i>	C513A	SM	■	■				
1472358	<i>rrs</i>	C513G	SM	■	■				
1472358	<i>rrs</i>	C513X	SM	■	■				
1472358	<i>rrs</i>	C513T	SM	■	■				
1472359	<i>rrs</i>	A514X	SM	■	■				
1472359	<i>rrs</i>	A514G	SM	■	■				
1472359	<i>rrs</i>	A514C	AMK	■	■				
1472359	<i>rrs</i>	A514T	SM	■	■				
1472359	<i>rrs</i>	A514C	SM	■	■				
1472360	<i>rrs</i>	A514G	SM	■	■				

D Mutationslisten

1472360	<i>rrs</i>	G515X	SM					
1472361	<i>rrs</i>	C516X	SM					
1472361	<i>rrs</i>	C516T	SM					
1472362	<i>rrs</i>	C517A	SM					
1472362	<i>rrs</i>	C517G	SM					
1472362	<i>rrs</i>	C517X	SM					
1472362	<i>rrs</i>	C517T	AMK					
1472362	<i>rrs</i>	C517T	SM					
1472473	<i>rrs</i>	G628C	SM					
1472644	<i>rrs</i>	C799T	SM					
1472733	<i>rrs</i>	G888A	SM					
1472750	<i>rrs</i>	C905A	SM					
1472750	<i>rrs</i>	C905G	SM					
1472751	<i>rrs</i>	A906G	SM					
1472752	<i>rrs</i>	A907G	SM					
1472752	<i>rrs</i>	A907T	SM					
1472752	<i>rrs</i>	A907C	SM					
1472753	<i>rrs</i>	A907G	SM					
1472753	<i>rrs</i>	A908G	SM					
1473003	<i>rrs</i>	G1158T	AMK					
1473074	<i>rrs</i>	C1229T	SM					
1473084	<i>rrs</i>	T1239C	SM					
1473183	<i>rrs</i>	A1338C	AMK					
1473246	<i>rrs</i>	A1401C	KAN					
1473246	<i>rrs</i>	A1401C	AMK					
1473246	<i>rrs</i>	A1401T	KAN					
1473246	<i>rrs</i>	A1401T	AMK					
1473246	<i>rrs</i>	A1401X	KAN					
1473246	<i>rrs</i>	A1401X	CPR					
1473246	<i>rrs</i>	A1401X	AMK					
1473246	<i>rrs</i>	A1402G	KAN					
1473246	<i>rrs</i>	A1402G	AMK					
1473246	<i>rrs</i>	A1401G	AMK					
1473246	<i>rrs</i>	A1401G	CPR					
1473246	<i>rrs</i>	A1401G	KAN					

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

1473246	<i>rrs</i>	A1401G	SM				
1473247	<i>rrs</i>	C1402X	KAN	■			
1473247	<i>rrs</i>	C1402X	CPR				
1473247	<i>rrs</i>	C1402X	AMK	■			
1473247	<i>rrs</i>	C1402A	AMK				
1473247	<i>rrs</i>	C1402T	AMK	■			
1473247	<i>rrs</i>	C1402A	CPR				
1473247	<i>rrs</i>	C1402T	CPR				
1473247	<i>rrs</i>	C1402A	KAN	■			
1473247	<i>rrs</i>	C1402T	KAN	■			
1473247	<i>rrs</i>	C1402T	SM				
1473329	<i>rrs</i>	G1484X	KAN				
1473329	<i>rrs</i>	G1484X	CPR				
1473329	<i>rrs</i>	G1484X	AMK				
1473329	<i>rrs</i>	G1484T	AMK				
1473329	<i>rrs</i>	G1484T	CPR				
1473329	<i>rrs</i>	G1484T	KAN				
1473329	<i>rrs</i>	G1484T	SM				
1475956	<i>rrl</i>	G2299T	LZD				
1476471	<i>rrl</i>	G2814T	LZD				
1673348	<i>inhA</i> P	A-92T	INH				
1673373	<i>inhA</i> P	G-67C	INH				
1673416	<i>inhA</i> P	G-24T	INH				
1673423	<i>inhA</i> P	G-17T	ETH				
1673423	<i>inhA</i> P	G-17T	INH	■			
1673424	<i>inhA</i> P	A-16G	INH				
1673425	<i>inhA</i> P	C-15T	ETH				
1673425	<i>inhA</i> P	C-15T	INH	■			
1673429	<i>inhA</i> P	A-11T	INH				
1673432	<i>inhA</i> P	T-8C	ETH				
1673432	<i>inhA</i> P	T-8A	INH	■			
1673432	<i>inhA</i> P	T-8C	INH				
1673432	<i>inhA</i> P	T-8G	INH				
1673435	<i>inhA</i> P	T-5A	INH				
1674225	<i>inhA</i>	K8N	INH				

D Mutationslisten

1674232	<i>inhA</i>	L11V	INH	
1674248	<i>inhA</i>	I16T	INH	
1674262	<i>inhA</i>	I21V	INH	
1674263	<i>inhA</i>	I21T	ETH	
1674263	<i>inhA</i>	I21N	INH	
1674263	<i>inhA</i>	I21T	INH	
1674319	<i>inhA</i>	G40W	INH	
1674341	<i>inhA</i>	I47T	INH	
1674481	<i>inhA</i>	S94A	ETH	
1674481	<i>inhA</i>	S94A	INH	
1674482	<i>inhA</i>	S94L	INH	
1674482	<i>inhA</i>	S94W	INH	
1674485	<i>inhA</i>	I95T	INH	
1674782	<i>inhA</i>	I194T	ETH	
1674782	<i>inhA</i>	I194T	INH	
1833909	<i>rpsA</i>	D123A	PZA	
1834325	<i>rpsA</i>	V262M	PZA	
1834840	<i>rpsA</i>	E433D	PZA	
1834852	<i>rpsA</i>	CGCC- 1312C	PZA	
1834961	<i>rpsA</i>	R474W	PZA	
1834962	<i>rpsA</i>	R474L	PZA	
1917946	<i>tlyA</i>	R3Stop	CPR	
1917961	<i>tlyA</i>	GA22G	CPR	
1917964	<i>tlyA</i>	GC25G	CPR	
1917979	<i>tlyA</i>	R14W	CPR	
1917991	<i>tlyA</i>	R18Stop	CPR	
1918003	<i>tlyA</i>	Q22Stop	CPR	
1918005	<i>tlyA</i>	GGCC- 66G	CPR	
1918139	<i>tlyA</i>	A67E	CPR	
1918144	<i>tlyA</i>	K69E	CPR	
1918211	<i>tlyA</i>	A91E	CPR	
1918248	<i>tlyA</i>	TG309T	CPR	
1918292	<i>tlyA</i>	L118P	CPR	

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

1918322	<i>tlyA</i>	V128E	CPR	
1918338	<i>tlyA</i>	GA399G	CPR	
1918388	<i>tlyA</i>	L150P	CPR	
1918415	<i>tlyA</i>	CG476C	CPR	
1918487	<i>tlyA</i>	P183L	CPR	
1918489	<i>tlyA</i>	Q184K	CPR	
1918489	<i>tlyA</i>	Q184Stop	CPR	
1918492	<i>tlyA</i>	F185L	CPR	
1918494	<i>tlyA</i>	F185L	CPR	
1918524	<i>tlyA</i>	TG585T	CPR	
1918591	<i>tlyA</i>	CT652C	CPR	
1918611	<i>tlyA</i>	CGT672C	CPR	
1918647	<i>tlyA</i>	N236K	CPR	
1918651	<i>tlyA</i>	E238K	CPR	
1918696	<i>tlyA</i>	GC757G	CPR	
2102240	<i>ndh</i>	R268H	INH	
			ETH	
2102715	<i>ndh</i>	T110A	INH	
			ETH	
2153906	<i>katG</i>	TGTC- 2206T	INH	
2153908	<i>katG</i>	D735A	INH	
2153909	<i>katG</i>	D735N	INH	
2153928	<i>katG</i>	W728C	INH	
2153932	<i>katG</i>	A727D	INH	
2153962	<i>katG</i>	Q717P	INH	
2153966	<i>katG</i>	A716P	INH	
2153975	<i>katG</i>	A713P	INH	
2153983	<i>katG</i>	V710A	INH	
2154014	<i>katG</i>	S700P	INH	
2154028	<i>katG</i>	D695A	INH	
2154059	<i>katG</i>	G685R	INH	
2154154	<i>katG</i>	L653P	INH	
2154205	<i>katG</i>	A636E	INH	
2154212	<i>katG</i>	L634F	INH	

D Mutationslisten

2154227	<i>katG</i>	G629S	INH
2154256	<i>katG</i>	L619P	INH
2154262	<i>katG</i>	A1850- AGT	INH
2154285	<i>katG</i>	M609I	INH
2154293	<i>katG</i>	E607K	INH
2154334	<i>katG</i>	G593D	INH
2154347	<i>katG</i>	P589T	INH
2154352	<i>katG</i>	L587P	INH
2154353	<i>katG</i>	L587M	INH
2154391	<i>katG</i>	A574V	INH
2154391	<i>katG</i>	A574E	INH
2154395	<i>katG</i>	D573N	INH
2154412	<i>katG</i>	F567S	INH
2154444	<i>katG</i>	TG1668T	INH
2154460	<i>katG</i>	A551G	INH
2154463	<i>katG</i>	A550D	INH
2154527	<i>katG</i>	N529D	INH
2154538	<i>katG</i>	Q525P	INH
2154569	<i>katG</i>	R515C	INH
2154598	<i>katG</i>	W505S	INH
2154599	<i>katG</i>	W505R	INH
2154606	<i>katG</i>	Q502H	INH
2154625	<i>katG</i>	R496L	INH
2154641	<i>katG</i>	G491C	INH
2154648	<i>katG</i>	K488N	INH
2154658	<i>katG</i>	G485V	INH
2154662	<i>katG</i>	R484S	INH
2154681	<i>katG</i>	W477Stop	INH
2154682	<i>katG</i>	W477Stop	INH
2154695	<i>katG</i>	V473F	INH
2154722	<i>katG</i>	A464S	INH
2154725	<i>katG</i>	R463W	INH
2154727	<i>katG</i>	I462T	INH
2154730	<i>katG</i>	Q461P	INH

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

2154737	<i>katG</i>	K459Stop	INH		
2154742	<i>katG</i>	S457I	INH		
2154767	<i>katG</i>	L449F	INH		
2154782	<i>katG</i>	A444T	INH		
2154782	<i>katG</i>	C1330CT	INH		
2154808	<i>katG</i>	T435I	INH		
2154812	<i>katG</i>	Q434Stop	INH		
2154827	<i>katG</i>	P429S	INH		
2154830	<i>katG</i>	G428R	INH		
2154841	<i>katG</i>	A424V	INH		
2154841	<i>katG</i>	A424E	INH		
2154841	<i>katG</i>	A424G	INH		
2154857	<i>katG</i>	D419H	INH		
2154859	<i>katG</i>	R418L	INH		
2154859	<i>katG</i>	R418Q	INH		
2154870	<i>katG</i>	K414N	INH		
2154875	<i>katG</i>	Y413H	INH		
2154886	<i>katG</i>	A409D	INH		
2154908	<i>katG</i>	CG1204C	INH		
2154921	<i>katG</i>	W397Stop	INH		
2154922	<i>katG</i>	W397Stop	INH		
2154932	<i>katG</i>	T394A	INH		
2154934	<i>katG</i>	I393N	INH		
2154941	<i>katG</i>	C1171C- TAG	INH		
2154961	<i>katG</i>	L384R	INH		
2154970	<i>katG</i>	D381G	INH		
2154973	<i>katG</i>	T380I	INH		
2154976	<i>katG</i>	A379V	INH		
2154979	<i>katG</i>	L378P	INH		
2155018	<i>katG</i>	P365L	INH		
2155020	<i>katG</i>	GA1092G	INH		
2155030	<i>katG</i>	A361D	INH		
2155043	<i>katG</i>	D357N	INH		
2155043	<i>katG</i>	D357H	INH		

D Mutationslisten

2155058	<i>katG</i>	Q352Stop	INH			
2155064	<i>katG</i>	A350S	INH			
2155064	<i>katG</i>	A350T	INH			
2155074	<i>katG</i>	GC1038G	INH			
2155078	<i>katG</i>	K345T	INH			
2155090	<i>katG</i>	W341S	INH			
2155102	<i>katG</i>	Y337F	INH			
2155102	<i>katG</i>	Y337C	INH			
2155105	<i>katG</i>	L336R	INH			
2155108	<i>katG</i>	I335T	INH			
2155109	<i>katG</i>	I335V	INH			
2155121	<i>katG</i>	S331C	INH			
2155128	<i>katG</i>	W328C	INH			
2155129	<i>katG</i>	W328L	INH			
2155129	<i>katG</i>	W328S	INH			
2155130	<i>katG</i>	W328G	INH			
2155135	<i>katG</i>	T326M	INH			
2155142	<i>katG</i>	T324P	INH			
2155149	<i>katG</i>	W321C	INH			
2155149	<i>katG</i>	W321Stop	INH			
2155150	<i>katG</i>	W321L	INH			
2155150	<i>katG</i>	W321Stop	INH			
2155150	<i>katG</i>	W321S	INH			
2155151	<i>katG</i>	W321R	INH			
2155163	<i>katG</i>	I317V	INH			
2155163	<i>katG</i>	I317L	INH			
2155165	<i>katG</i>	G316D	INH			
2155166	<i>katG</i>	G316S	INH			
2155167	<i>katG</i>	S315X	INH			
2155167	<i>katG</i>	S315R	INH			
2155168	<i>katG</i>	S315X	INH			
2155168	<i>katG</i>	S315I	INH			
2155168	<i>katG</i>	S315N	INH			
2155168	<i>katG</i>	S315T	INH			
2155169	<i>katG</i>	S315X	INH			

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

2155169	<i>katG</i>	S315G	INH			
2155169	<i>katG</i>	S315R	INH			
2155179	<i>katG</i>	D311E	INH			
2155180	<i>katG</i>	D311G	INH			
2155181	<i>katG</i>	D311Y	INH			
2155183	<i>katG</i>	TT929T	INH			
2155186	<i>katG</i>	G309D	INH			
2155187	<i>katG</i>	G309C	INH			
2155187	<i>katG</i>	G309S	INH			
2155190	<i>katG</i>	T308P	INH			
2155192	<i>katG</i>	G307E	INH			
2155192	<i>katG</i>	G307A	INH			
2155193	<i>katG</i>	G307R	INH			
2155198	<i>katG</i>	G305A	INH			
2155201	<i>katG</i>	Y304S	INH			
2155206	<i>katG</i>	S302R	INH			
2155208	<i>katG</i>	S302R	INH			
2155210	<i>katG</i>	T902TA	INH			
2155212	<i>katG</i>	W300C	INH			
2155212	<i>katG</i>	W300Stop	INH			
2155213	<i>katG</i>	W300Stop	INH			
2155214	<i>katG</i>	W300G	INH			
2155216	<i>katG</i>	G299A	INH			
2155217	<i>katG</i>	G299C	INH			
2155222	<i>katG</i>	G297V	INH			
2155228	<i>katG</i>	Q295P	INH			
2155241	<i>katG</i>	A291P	INH			
2155245	<i>katG</i>	E289D	INH			
2155258	<i>katG</i>	G285D	INH			
2155270	<i>katG</i>	A281V	INH			
2155276	<i>katG</i>	G279D	INH			
2155288	<i>katG</i>	T275S	INH			
2155289	<i>katG</i>	T275S	INH			
2155289	<i>katG</i>	T275A	INH			
2155289	<i>katG</i>	T275P	INH			

D Mutationslisten

2155308	<i>katG</i>	GC804G	INH
2155316	<i>katG</i>	TC796T	INH
2155322	<i>katG</i>	A264T	INH
2155327	<i>katG</i>	T262R	INH
2155331	<i>katG</i>	E261K	INH
2155335	<i>katG</i>	D259E	INH
2155339	<i>katG</i>	N258S	INH
2155341	<i>katG</i>	M257I	INH
2155342	<i>katG</i>	M257T	INH
2155356	<i>katG</i>	F252L	INH
2155358	<i>katG</i>	F252L	INH
2155360	<i>katG</i>	T251M	INH
2155367	<i>katG</i>	R249C	INH
2155385	<i>katG</i>	A243S	INH
2155389	<i>katG</i>	P241P	INH
2155398	<i>katG</i>	N238K	INH
2155405	<i>katG</i>	N236T	INH
2155411	<i>katG</i>	G234E	INH
2155412	<i>katG</i>	G234R	INH
2155423	<i>katG</i>	V230A	INH
2155426	<i>katG</i>	Y229F	INH
2155442	<i>katG</i>	Q224E	INH
2155458	<i>katG</i>	N218K	INH
2155514	<i>katG</i>	K200Stop	INH
2155518	<i>katG</i>	W198Stop	INH
2155519	<i>katG</i>	W198Stop	INH
2155529	<i>katG</i>	E195K	INH
2155539	<i>katG</i>	CC573C	INH
2155541	<i>katG</i>	W191R	INH
2155550	<i>katG</i>	CC562C	INH
2155555	<i>katG</i>	G186V	INH
2155573	<i>katG</i>	T180K	INH
2155584	<i>katG</i>	M176I	INH
2155597	<i>katG</i>	A172V	INH
2155598	<i>katG</i>	A172T	INH

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

2155606	<i>katG</i>	G169A	INH
2155628	<i>katG</i>	A162T	INH
2155633	<i>katG</i>	S160L	INH
2155648	<i>katG</i>	Y155C	INH
2155648	<i>katG</i>	Y155S	INH
2155669	<i>katG</i>	L148R	INH
2155684	<i>katG</i>	K143T	INH
2155687	<i>katG</i>	D142A	INH
2155689	<i>katG</i>	L141F	INH
2155692	<i>katG</i>	S140R	INH
2155693	<i>katG</i>	S140N	INH
2155694	<i>katG</i>	S140G	INH
2155694	<i>katG</i>	S140R	INH
2155697	<i>katG</i>	A139P	INH
2155699	<i>katG</i>	N138S	INH
2155699	<i>katG</i>	N138T	INH
2155700	<i>katG</i>	N138D	INH
2155700	<i>katG</i>	N138H	INH
2155729	<i>katG</i>	R128P	INH
2155732	<i>katG</i>	Q127P	INH
2155734	<i>katG</i>	M126I	INH
2155738	<i>katG</i>	C374CA- TG	INH
2155739	<i>katG</i>	G125C	INH
2155739	<i>katG</i>	C373CC	INH
2155742	<i>katG</i>	CCCCGG- CGCCGC- C370C	INH
2155750	<i>katG</i>	G121V	INH
2155751	<i>katG</i>	G121C	INH
2155762	<i>katG</i>	D117A	INH
2155783	<i>katG</i>	A110V	INH
2155786	<i>katG</i>	A109V	INH
2155788	<i>katG</i>	H108Q	INH
2155790	<i>katG</i>	H108D	INH

D Mutationslisten

2155791	<i>katG</i>	W107Stop	INH
2155792	<i>katG</i>	W107Stop	INH
2155793	<i>katG</i>	W107R	INH
2155795	<i>katG</i>	A106V	INH
2155801	<i>katG</i>	R104L	INH
2155801	<i>katG</i>	R104Q	INH
2155816	<i>katG</i>	G99E	INH
2155826	<i>katG</i>	G96C	INH
2155831	<i>katG</i>	D94G	INH
2155831	<i>katG</i>	D94A	INH
2155841	<i>katG</i>	W91R	INH
2155842	<i>katG</i>	W90Stop	INH
2155843	<i>katG</i>	W90Stop	INH
2155849	<i>katG</i>	Q88R	INH
2155859	<i>katG</i>	T85P	INH
2155860	<i>katG</i>	M84I	INH
2155891	<i>katG</i>	D74G	INH
2155892	<i>katG</i>	D74Y	INH
2155897	<i>katG</i>	D72G	INH
2155900	<i>katG</i>	I71N	INH
2155916	<i>katG</i>	A66P	INH
2155916	<i>katG</i>	C196- CGGGG	INH
2155919	<i>katG</i>	A65T	INH
2155923	<i>katG</i>	D63E	INH
2155926	<i>katG</i>	G186GG	INH
2155931	<i>katG</i>	A61T	INH
2155945	<i>katG</i>	TCAGCG- ACGGCC- GGGT- 167T	INH
2155962	<i>katG</i>	TT150T	INH
2155969	<i>katG</i>	L48Q	INH
2155998	<i>katG</i>	W38Stop	INH
2155999	<i>katG</i>	W38Stop	INH

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

2156002	<i>katG</i>	TC110T	INH	
2156009	<i>katG</i>	N35D	INH	
2156013	<i>katG</i>	G99GT	INH	
2156056	<i>katG</i>	G19D	INH	
2156062	<i>katG</i>	S17N	INH	
2156076	<i>katG</i>	GGTGGT- TTC36G	INH	
2156078	<i>katG</i>	T12P	INH	
2156081	<i>katG</i>	T11A	INH	
2156108	<i>katG</i>	P2S	INH	
2156110	<i>katG</i>	V1A	INH	
2156111	<i>katG</i>	V1L	INH	
2156118	<i>katG</i> P	C-7T	INH	
2156121	<i>katG</i> P	T-10G	INH	
2156123	<i>katG</i> P	C-12T	INH	
2288683	<i>pncA</i>	Stop187R	PZA	
2288688	<i>pncA</i>	S185I	PZA	
2288688	<i>pncA</i>	S185T	PZA	
2288697	<i>pncA</i>	L182S	PZA	
2288699	<i>pncA</i>	E181D	PZA	
2288703	<i>pncA</i>	V180A	PZA	
2288703	<i>pncA</i>	V180G	PZA	
2288704	<i>pncA</i>	V180F	PZA	
2288705	<i>pncA</i>	S179R	PZA	
2288707	<i>pncA</i>	S179R	PZA	
2288709	<i>pncA</i>	G533GG	PZA	
2288717	<i>pncA</i>	M175I	PZA	
2288718	<i>pncA</i>	M175T	PZA	
2288718	<i>pncA</i>	M175R	PZA	
2288719	<i>pncA</i>	M175V	PZA	
2288721	<i>pncA</i>	E174G	PZA	
2288723	<i>pncA</i>	C519CN	PZA	
2288727	<i>pncA</i>	L172R	PZA	
2288727	<i>pncA</i>	L172P	PZA	
2288729	<i>pncA</i>	CG513C	PZA	

D Mutationslisten

2288730	<i>pncA</i>	A171V	PZA	
2288730	<i>pncA</i>	A171E	PZA	
2288731	<i>pncA</i>	A171T	PZA	
2288731	<i>pncA</i>	A171P	PZA	
2288739	<i>pncA</i>	T168N	PZA	
2288740	<i>pncA</i>	T168P	PZA	
2288740	<i>pncA</i>	T502TGC	PZA	
2288749	<i>pncA</i>	A165T	PZA	
2288752	<i>pncA</i>	S164P	PZA	
2288754	<i>pncA</i>	V163A	PZA	
2288757	<i>pncA</i>	G162D	PZA	
2288761	<i>pncA</i>	A161P	PZA	
2288761	<i>pncA</i>	C481- CGTCA	PZA	
2288764	<i>pncA</i>	T160P	PZA	
2288766	<i>pncA</i>	L159R	PZA	
2288766	<i>pncA</i>	L159P	PZA	
2288766	<i>pncA</i>	AG476A	PZA	
2288772	<i>pncA</i>	V157A	PZA	
2288772	<i>pncA</i>	V157G	PZA	
2288776	<i>pncA</i>	G466GA	PZA	
2288778	<i>pncA</i>	V155G	PZA	
2288778	<i>pncA</i>	V155A	PZA	
2288779	<i>pncA</i>	V155M	PZA	
2288779	<i>pncA</i>	V155L	PZA	
2288781	<i>pncA</i>	R154M	PZA	
2288782	<i>pncA</i>	R154G	PZA	
2288784	<i>pncA</i>	T153N	PZA	
2288788	<i>pncA</i>	CCAAGC- CAT454 C	PZA	
2288789	<i>pncA</i>	CA453C	PZA	
2288798	<i>pncA</i>	GC444G	PZA	
2288800	<i>pncA</i>	R148S	PZA	
2288805	<i>pncA</i>	A146E	PZA	

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

2288805	<i>pncA</i>	A146V	PZA	
2288805	<i>pncA</i>	GC437G	PZA	
2288806	<i>pncA</i>	A146P	PZA	
2288806	<i>pncA</i>	A146T	PZA	
2288813	<i>pncA</i>	G429GCC	PZA	
2288815	<i>pncA</i>	A143T	PZA	
2288815	<i>pncA</i>	A143P	PZA	
2288817	<i>pncA</i>	T142M	PZA	
2288817	<i>pncA</i>	T142K	PZA	
2288818	<i>pncA</i>	T142A	PZA	
2288818	<i>pncA</i>	T142P	PZA	
2288820	<i>pncA</i>	Q141P	PZA	
2288821	<i>pncA</i>	Q141Stop	PZA	
2288821	<i>pncA</i>	G421GCC	PZA	
2288821	<i>pncA</i>	G421GC	PZA	
2288823	<i>pncA</i>	R140P	PZA	
2288823	<i>pncA</i>	R140H	PZA	
2288824	<i>pncA</i>	GCA418G	PZA	
2288824	<i>pncA</i>	R140S	PZA	
2288824	<i>pncA</i>	R140G	PZA	
2288826	<i>pncA</i>	V139G	PZA	
2288826	<i>pncA</i>	V139A	PZA	
2288827	<i>pncA</i>	V139L	PZA	
2288827	<i>pncA</i>	C415CC	PZA	
2288827	<i>pncA</i>	V139M	PZA	
2288828	<i>pncA</i>	C138W	PZA	
2288828	<i>pncA</i>	C138Stop	PZA	
2288829	<i>pncA</i>	C138Y	PZA	
2288829	<i>pncA</i>	C138S	PZA	
2288830	<i>pncA</i>	C138R	PZA	
2288830	<i>pncA</i>	C138S	PZA	
2288832	<i>pncA</i>	H137R	PZA	
2288832	<i>pncA</i>	H137P	PZA	
2288833	<i>pncA</i>	H137D	PZA	
2288834	<i>pncA</i>	A408AG	PZA	

D Mutationslisten

2288835	<i>pncA</i>	TC407T	PZA	
2288835	<i>pncA</i>	D136G	PZA	
2288836	<i>pncA</i>	D136Y	PZA	
2288836	<i>pncA</i>	D136N	PZA	
2288836	<i>pncA</i>	D136H	PZA	
2288836	<i>pncA</i>	CGGTGG- CAATAC- 406C	PZA	
2288838	<i>pncA</i>	T135N	PZA	
2288838	<i>pncA</i>	G404GG	PZA	
2288839	<i>pncA</i>	T135P	PZA	
2288841	<i>pncA</i>	A134V	PZA	
2288842	<i>pncA</i>	A134S	PZA	
2288844	<i>pncA</i>	A398AA	PZA	
2288844	<i>pncA</i>	I133N	PZA	
2288844	<i>pncA</i>	I133T	PZA	
2288847	<i>pncA</i>	G132V	PZA	
2288847	<i>pncA</i>	G132D	PZA	
2288847	<i>pncA</i>	G132A	PZA	
2288848	<i>pncA</i>	G132C	PZA	
2288848	<i>pncA</i>	G132S	PZA	
2288849	<i>pncA</i>	G393GC	PZA	
2288850	<i>pncA</i>	AC392A	PZA	
2288850	<i>pncA</i>	A392ACC	PZA	
2288851	<i>pncA</i>	V131F	PZA	
2288852	<i>pncA</i>	CACATC- GACCTC- 390C	PZA	
2288852	<i>pncA</i>	CACAT- 390C	PZA	
2288853	<i>pncA</i>	V130G	PZA	
2288853	<i>pncA</i>	V130A	PZA	
2288853	<i>pncA</i>	A389AA- TCGACC- TC	PZA	

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

2288853	<i>pncA</i>	A389AC- ATCGAC- CT	PZA	
2288857	<i>pncA</i>	D129Y	PZA	
2288859	<i>pncA</i>	A383ACT	PZA	
2288859	<i>pncA</i>	V128G	PZA	
2288859	<i>pncA</i>	ACC383A	PZA	
2288868	<i>pncA</i>	V125D	PZA	
2288868	<i>pncA</i>	V125G	PZA	
2288869	<i>pncA</i>	V125F	PZA	
2288873	<i>pncA</i>	G369GN	PZA	
2288873	<i>pncA</i>	G369GCT	PZA	
2288874	<i>pncA</i>	R123P	PZA	
2288880	<i>pncA</i>	R121P	PZA	
2288883	<i>pncA</i>	L120R	PZA	
2288883	<i>pncA</i>	L120Q	PZA	
2288883	<i>pncA</i>	L120P	PZA	
2288885	<i>pncA</i>	W119C	PZA	
2288885	<i>pncA</i>	W119Stop	PZA	
2288886	<i>pncA</i>	W119S	PZA	
2288886	<i>pncA</i>	W119Stop	PZA	
2288887	<i>pncA</i>	W119G	PZA	
2288887	<i>pncA</i>	W119R	PZA	
2288889	<i>pncA</i>	N118T	PZA	
2288895	<i>pncA</i>	L116R	PZA	
2288895	<i>pncA</i>	L116P	PZA	
2288896	<i>pncA</i>	L116V	PZA	
2288900	<i>pncA</i>	CG342C	PZA	
2288902	<i>pncA</i>	T114A	PZA	
2288902	<i>pncA</i>	T114P	PZA	
2288920	<i>pncA</i>	G108R	PZA	
2288923	<i>pncA</i>	E107Stop	PZA	
2288928	<i>pncA</i>	G105D	PZA	
2288929	<i>pncA</i>	CG313C	PZA	
2288930	<i>pncA</i>	S104R	PZA	

D Mutationslisten

2288930	<i>pncA</i>	GCTGTA 312 G	PZA	
2288931	<i>pncA</i>	S104I	PZA	
2288932	<i>pncA</i>	S104R	PZA	
2288933	<i>pncA</i>	Y103Stop	PZA	
2288934	<i>pncA</i>	Y103C	PZA	
2288934	<i>pncA</i>	Y103S	PZA	
2288935	<i>pncA</i>	Y103D	PZA	
2288935	<i>pncA</i>	A307AG	PZA	
2288935	<i>pncA</i>	Y103H	PZA	
2288937	<i>pncA</i>	A102V	PZA	
2288938	<i>pncA</i>	A102T	PZA	
2288940	<i>pncA</i>	C302CC	PZA	
2288940	<i>pncA</i>	C302CT	PZA	
2288944	<i>pncA</i>	T100A	PZA	
2288944	<i>pncA</i>	T100P	PZA	
2288945	<i>pncA</i>	Y99Stop	PZA	
2288952	<i>pncA</i>	G97A	PZA	
2288952	<i>pncA</i>	G97D	PZA	
2288953	<i>pncA</i>	CC289C	PZA	
2288953	<i>pncA</i>	G97S	PZA	
2288954	<i>pncA</i>	K96N	PZA	
2288954	<i>pncA</i>	C288CA	PZA	
2288955	<i>pncA</i>	K96R	PZA	
2288955	<i>pncA</i>	K96T	PZA	
2288956	<i>pncA</i>	K96E	PZA	
2288956	<i>pncA</i>	K96Q	PZA	
2288957	<i>pncA</i>	Y95Stop	PZA	
2288960	<i>pncA</i>	F94L	PZA	
2288961	<i>pncA</i>	F94C	PZA	
2288961	<i>pncA</i>	F94S	PZA	
2288962	<i>pncA</i>	F94L	PZA	
2288971	<i>pncA</i>	E91Stop	PZA	
2288973	<i>pncA</i>	I90S	PZA	
2288979	<i>pncA</i>	S88Stop	PZA	

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

2288982	<i>pncA</i>	T87M	PZA	
2288988	<i>pncA</i>	L85R	PZA	
2288988	<i>pncA</i>	L85P	PZA	
2288997	<i>pncA</i>	H82L	PZA	
2288997	<i>pncA</i>	H82R	PZA	
2288998	<i>pncA</i>	H82Y	PZA	
2288998	<i>pncA</i>	H82D	PZA	
2289000	<i>pncA</i>	F81C	PZA	
2289000	<i>pncA</i>	F81S	PZA	
2289001	<i>pncA</i>	F81V	PZA	
2289002	<i>pncA</i>	GT240G	PZA	
2289006	<i>pncA</i>	A79G	PZA	
2289009	<i>pncA</i>	G78D	PZA	
2289010	<i>pncA</i>	G78S	PZA	
2289015	<i>pncA</i>	T76I	PZA	
2289016	<i>pncA</i>	T76P	PZA	
2289020	<i>pncA</i>	G222GC	PZA	
2289023	<i>pncA</i>	G219GC- GGCAAT- GCG	PZA	
2289026	<i>pncA</i>	C72Stop	PZA	
2289026	<i>pncA</i>	C72W	PZA	
2289027	<i>pncA</i>	C72Y	PZA	
2289028	<i>pncA</i>	C72R	PZA	
2289029	<i>pncA</i>	H71Q	PZA	
2289030	<i>pncA</i>	H71R	PZA	
2289031	<i>pncA</i>	H71Y	PZA	
2289031	<i>pncA</i>	H71N	PZA	
2289031	<i>pncA</i>	H71D	PZA	
2289036	<i>pncA</i>	P69L	PZA	
2289036	<i>pncA</i>	P69R	PZA	
2289038	<i>pncA</i>	W68C	PZA	
2289038	<i>pncA</i>	W68Stop	PZA	
2289039	<i>pncA</i>	W68L	PZA	
2289039	<i>pncA</i>	W68Stop	PZA	

D Mutationslisten

2289039	<i>pncA</i>	W68S	PZA	
2289040	<i>pncA</i>	W68R	PZA	
2289040	<i>pncA</i>	W68G	PZA	
2289043	<i>pncA</i>	S67P	PZA	
2289046	<i>pncA</i>	S66P	PZA	
2289049	<i>pncA</i>	A193AT	PZA	
2289050	<i>pncA</i>	Y64Stop	PZA	
2289052	<i>pncA</i>	Y64D	PZA	
2289054	<i>pncA</i>	D63G	PZA	
2289054	<i>pncA</i>	D63A	PZA	
2289055	<i>pncA</i>	D63H	PZA	
2289057	<i>pncA</i>	P62L	PZA	
2289057	<i>pncA</i>	P62R	PZA	
2289061	<i>pncA</i>	T61P	PZA	
2289067	<i>pncA</i>	S59P	PZA	
2289068	<i>pncA</i>	F58L	PZA	
2289069	<i>pncA</i>	F58S	PZA	
2289069	<i>pncA</i>	AA173A	PZA	
2289070	<i>pncA</i>	F58L	PZA	
2289071	<i>pncA</i>	H57Q	PZA	
2289072	<i>pncA</i>	H57L	PZA	
2289072	<i>pncA</i>	H57R	PZA	
2289072	<i>pncA</i>	H57P	PZA	
2289073	<i>pncA</i>	H57D	PZA	
2289073	<i>pncA</i>	H57Y	PZA	
2289073	<i>pncA</i>	H57N	PZA	
2289079	<i>pncA</i>	CC163C	PZA	
2289079	<i>pncA</i>	C163CA	PZA	
2289080	<i>pncA</i>	CG162C	PZA	
2289081	<i>pncA</i>	P54R	PZA	
2289081	<i>pncA</i>	P54L	PZA	
2289081	<i>pncA</i>	P54Q	PZA	
2289082	<i>pncA</i>	P54S	PZA	
2289082	<i>pncA</i>	P54T	PZA	
2289083	<i>pncA</i>	GT159G	PZA	

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

2289084	<i>pncA</i>	D53A	PZA	
2289085	<i>pncA</i>	D53N	PZA	
2289089	<i>pncA</i>	H51Q	PZA	
2289090	<i>pncA</i>	H51R	PZA	
2289090	<i>pncA</i>	H51P	PZA	
2289091	<i>pncA</i>	H51Y	PZA	
2289091	<i>pncA</i>	H51N	PZA	
2289096	<i>pncA</i>	D49V	PZA	
2289096	<i>pncA</i>	D49G	PZA	
2289096	<i>pncA</i>	D49A	PZA	
2289097	<i>pncA</i>	D49N	PZA	
2289097	<i>pncA</i>	D49H	PZA	
2289100	<i>pncA</i>	K48E	PZA	
2289100	<i>pncA</i>	K48Stop	PZA	
2289102	<i>pncA</i>	T47S	PZA	
2289103	<i>pncA</i>	T47S	PZA	
2289103	<i>pncA</i>	T47A	PZA	
2289103	<i>pncA</i>	T47P	PZA	
2289105	<i>pncA</i>	A46V	PZA	
2289105	<i>pncA</i>	A46E	PZA	
2289105	<i>pncA</i>	G137GC	PZA	
2289106	<i>pncA</i>	A46P	PZA	
2289108	<i>pncA</i>	V45G	PZA	
2289111	<i>pncA</i>	V44G	PZA	
2289114	<i>pncA</i>	H43P	PZA	
2289119	<i>pncA</i>	Y41Stop	PZA	
2289121	<i>pncA</i>	Y41H	PZA	
2289130	<i>pncA</i>	CTTCGG- CCAGGT- AGTCGC- TGATGG- C112C	PZA	
2289133	<i>pncA</i>	E37Stop	PZA	
2289138	<i>pncA</i>	L35R	PZA	
2289138	<i>pncA</i>	L35P	PZA	

D Mutationslisten

2289140	<i>pncA</i>	Y34Stop	PZA	
2289141	<i>pncA</i>	T101TA	PZA	
2289141	<i>pncA</i>	Y34S	PZA	
2289142	<i>pncA</i>	Y34D	PZA	
2289150	<i>pncA</i>	I31S	PZA	
2289157	<i>pncA</i>	GG85G	PZA	
2289159	<i>pncA</i>	A28D	PZA	
2289162	<i>pncA</i>	L27P	PZA	
2289165	<i>pncA</i>	A26G	PZA	
2289170	<i>pncA</i>	GC72G	PZA	
2289171	<i>pncA</i>	CC71C	PZA	
2289171	<i>pncA</i>	G24D	PZA	
2289172	<i>pncA</i>	G24R	PZA	
2289174	<i>pncA</i>	G23V	PZA	
2289177	<i>pncA</i>	GT65G	PZA	
2289180	<i>pncA</i>	V21G	PZA	
2289182	<i>pncA</i>	CG60C	PZA	
2289186	<i>pncA</i>	L19R	PZA	
2289186	<i>pncA</i>	L19P	PZA	
2289189	<i>pncA</i>	G53GCC	PZA	
2289190	<i>pncA</i>	S18P	PZA	
2289192	<i>pncA</i>	G17V	PZA	
2289192	<i>pncA</i>	G17D	PZA	
2289193	<i>pncA</i>	G17S	PZA	
2289194	<i>pncA</i>	AC48A	PZA	
2289200	<i>pncA</i>	C14Stop	PZA	
2289200	<i>pncA</i>	C14W	PZA	
2289201	<i>pncA</i>	C14Y	PZA	
2289202	<i>pncA</i>	C14G	PZA	
2289202	<i>pncA</i>	C14R	PZA	
2289203	<i>pncA</i>	F13L	PZA	
2289204	<i>pncA</i>	F13S	PZA	
2289205	<i>pncA</i>	F13L	PZA	
2289206	<i>pncA</i>	D12E	PZA	
2289207	<i>pncA</i>	D12A	PZA	

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

2289208	<i>pncA</i>	D12N	PZA	
2289213	<i>pncA</i>	Q10R	PZA	
2289213	<i>pncA</i>	Q10P	PZA	
2289213	<i>pncA</i>	TG29T	PZA	
2289214	<i>pncA</i>	Q10Stop	PZA	
2289214	<i>pncA</i>	Q10K	PZA	
2289216	<i>pncA</i>	V9G	PZA	
2289216	<i>pncA</i>	V9A	PZA	
2289217	<i>pncA</i>	V9L	PZA	
2289218	<i>pncA</i>	D8E	PZA	
2289219	<i>pncA</i>	D8G	PZA	
2289219	<i>pncA</i>	D8A	PZA	
2289220	<i>pncA</i>	D8Y	PZA	
2289220	<i>pncA</i>	D8N	PZA	
2289222	<i>pncA</i>	V7D	PZA	
2289222	<i>pncA</i>	V7G	PZA	
2289223	<i>pncA</i>	V7F	PZA	
2289223	<i>pncA</i>	V7I	PZA	
2289225	<i>pncA</i>	I6T	PZA	
2289225	<i>pncA</i>	I6S	PZA	
2289226	<i>pncA</i>	T16TG	PZA	
2289226	<i>pncA</i>	I6L	PZA	
2289228	<i>pncA</i>	I5S	PZA	
2289230	<i>pncA</i>	CAACGC- CCGCAT- 12C	PZA	
2289231	<i>pncA</i>	L4W	PZA	
2289231	<i>pncA</i>	L4S	PZA	
2289233	<i>pncA</i>	CGCCCG- CAT9C	PZA	
2289234	<i>pncA</i>	A3E	PZA	
2289235	<i>pncA</i>	A3P	PZA	
2289239	<i>pncA</i>	M1I	PZA	
2289240	<i>pncA</i>	M1K	PZA	
2289240	<i>pncA</i>	M1T	PZA	

D Mutationslisten

2289248	<i>pncA</i> P	A-7G	PZA	
2289251	<i>pncA</i> P	A-10G	PZA	
2289252	<i>pncA</i> P	T-11C	PZA	
2289252	<i>pncA</i> P	T-11G	PZA	
2289253	<i>pncA</i> P	A-12C	PZA	
2289253	<i>pncA</i> P	A-12G	PZA	
2518310	<i>kasA</i>	D66N	INH	
2518345	<i>kasA</i>	M77I	INH	
2518476	<i>kasA</i>	R121K	INH	
2518919	<i>kasA</i>	G269S	INH	
2519048	<i>kasA</i>	G312S	INH	
2519274	<i>kasA</i>	G387D	INH	
2519351	<i>kasA</i>	F413L	INH	
2519353	<i>kasA</i>	F413L	INH	
2715340	<i>eis</i> P	G-8T	KAN	
2715342	<i>eis</i> P	C-10T	KAN	
2715342	<i>eis</i> P	C-10G	KAN	
2715344	<i>eis</i> P	G-12A	KAN	
2715345	<i>eis</i> P	T-13C	KAN	
2715346	<i>eis</i> P	G-14A	KAN	
2715347	<i>eis</i> P	G-15C	KAN	
2715369	<i>eis</i> P	C-37A	KAN	
2715385	<i>eis</i> P	A-53C	KAN	
2715386	<i>eis</i> P	C-54A	KAN	
2726112	<i>ahpC</i> P	C-81T	INH	
2726119	<i>ahpC</i> P	G-74A	INH	
2726127	<i>ahpC</i> P	G-66A	INH	
2726136	<i>ahpC</i> P	C-57T	INH	
2726139	<i>ahpC</i> P	C-54T	INH	
2726141	<i>ahpC</i> P	C-52T	INH	
2726142	<i>ahpC</i> P	G-51T	INH	
2726144	<i>ahpC</i> P	T-49G	INH	
2726145	<i>ahpC</i> P	G-48A	INH	
2726145	<i>ahpC</i> P	G-48T	INH	
2726149	<i>ahpC</i> P	T-44A	INH	

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

2726154	<i>ahpC</i> P	C-39T	INH			
2726173	<i>ahpC</i> P	C-20T	INH			
2726188	<i>ahpC</i> P	G-5A	INH			
2726189	<i>ahpC</i> P	A-4G	INH			
2726196	<i>ahpC</i>	P2S	INH			
2726206	<i>ahpC</i>	T5I	INH			
2726220	<i>ahpC</i>	F10I	INH			
2726289	<i>ahpC</i>	D33N	INH			
2726409	<i>ahpC</i>	D73H	INH			
2726418	<i>ahpC</i>	E76K	INH			
2726764	<i>ahpC</i>	L191R	INH			
3073808	<i>thyA</i>	R222G	PAS			
3073868	<i>thyA</i>	T202A	PAS			
3877949	<i>rpoA</i>	T187A	RMP			
3877949	<i>rpoA</i>	T187P	RMP			
3877960	<i>rpoA</i>	V183A	RMP			
3877960	<i>rpoA</i>	V183G	RMP			
4240423	<i>embC</i>	E187E	EMB			
4240592	<i>embC</i>	A244T	EMB			
4240601	<i>embC</i>	A247P	EMB			
4240614	<i>embC</i>	L251R	EMB			
4240623	<i>embC</i>	A254G	EMB			
4240676	<i>embC</i>	G272S	EMB			
4240715	<i>embC</i>	H285Y	EMB			
4240721	<i>embC</i>	V287F	EMB			
4240724	<i>embC</i>	G288W	EMB			
4240725	<i>embC</i>	G288V	EMB			
4240748	<i>embC</i>	Y296H	EMB			
4240749	<i>embC</i>	Y296S	EMB			
4240751	<i>embC</i>	I297L	EMB			
4240752	<i>embC</i>	I297T	EMB			
4240761	<i>embC</i>	M300R	EMB			
4240766	<i>embC</i>	R302G	EMB			
4240770	<i>embC</i>	V303G	EMB			
4240781	<i>embC</i>	A307T	EMB			

D Mutationslisten

4240785	<i>embC</i>	G308D	EMB		
4240787	<i>embC</i>	Y309N	EMB		
4240791	<i>embC</i>	M310K	EMB		
4240835	<i>embC</i>	G325S	EMB		
4240838	<i>embC</i>	W326R	EMB		
4240841	<i>embC</i>	Y327N	EMB		
4241078	<i>embC</i>	I406V	EMB		
4242198	<i>embC</i>	A779V	EMB		
4243217	<i>embA</i> P	C-16T	EMB		
4243221	<i>embA</i> P	C-12T	EMB		
4243222	<i>embA</i> P	C-11A	EMB		
4243225	<i>embA</i>	C-8A	EMB		
4243242	<i>embA</i>	D4N	EMB		
4243245	<i>embA</i>	G5S	EMB		
4243833	<i>embA</i>	A201T	EMB		
4244193	<i>embA</i>	G321S	EMB		
4244281	<i>embA</i>	G350D	EMB		
4244617	<i>embA</i>	A462V	EMB		
4245730	<i>embA</i>	D833A	EMB		
4246734	<i>embB</i>	L74R	EMB		
4247175	<i>embB</i>	A221G	EMB		
4247186	<i>embB</i>	A225S	EMB		
4247193	<i>embB</i>	V227G	EMB		
4247229	<i>embB</i>	L239P	EMB		
4247231	<i>embB</i>	D240H	EMB		
4247282	<i>embB</i>	R257W	EMB		
4247325	<i>embB</i>	A271V	EMB		
4247327	<i>embB</i>	S272C	EMB		
4247328	<i>embB</i>	S272I	EMB		
4247334	<i>embB</i>	R274P	EMB		
4247354	<i>embB</i>	A281P	EMB		
4247357	<i>embB</i>	V282L	EMB		
4247358	<i>embB</i>	V282G	EMB		
4247366	<i>embB</i>	F285L	EMB		
4247368	<i>embB</i>	F285L	EMB		

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

4247372	<i>embB</i>	F287V	EMB				
4247373	<i>embB</i>	F287C	EMB				
4247391	<i>embB</i>	I293N	EMB				
4247393	<i>embB</i>	G294S	EMB				
4247399	<i>embB</i>	N296H	EMB				
4247401	<i>embB</i>	N296K	EMB				
4247402	<i>embB</i>	S297A	EMB				
4247403	<i>embB</i>	S297L	EMB				
4247406	<i>embB</i>	S298W	EMB				
4247410	<i>embB</i>	D299E	EMB				
4247429	<i>embB</i>	M306X	EMB				
4247429	<i>embB</i>	M306L	EMB				
4247429	<i>embB</i>	M306L	EMB				
4247429	<i>embB</i>	M306V	EMB				
4247430	<i>embB</i>	M306X	EMB				
4247430	<i>embB</i>	M306T	EMB				
4247431	<i>embB</i>	M306X	EMB				
4247431	<i>embB</i>	M306I	EMB				
4247439	<i>embB</i>	V309G	EMB				
4247444	<i>embB</i>	D311H	EMB				
4247445	<i>embB</i>	D311G	EMB				
4247448	<i>embB</i>	H312R	EMB				
4247451	<i>embB</i>	A313V	EMB				
4247454	<i>embB</i>	G314A	EMB				
4247457	<i>embB</i>	Y315C	EMB				
4247462	<i>embB</i>	S317T	EMB				
4247465	<i>embB</i>	N318H	EMB				
4247466	<i>embB</i>	N318S	EMB				
4247467	<i>embB</i>	N318K	EMB				
4247468	<i>embB</i>	Y319N	EMB				
4247469	<i>embB</i>	Y319S	EMB				
4247469	<i>embB</i>	Y319C	EMB				
4247471	<i>embB</i>	F320L	EMB				
4247473	<i>embB</i>	F320L	EMB				
4247495	<i>embB</i>	GA982G	EMB				

D Mutationslisten

4247495	<i>embB</i>	D328Y	EMB	
4247496	<i>embB</i>	D328V	EMB	
4247496	<i>embB</i>	D328G	EMB	
4247501	<i>embB</i>	F330V	EMB	
4247504	<i>embB</i>	G331R	EMB	
4247507	<i>embB</i>	W332R	EMB	
4247508	<i>embB</i>	W332L	EMB	
4247510	<i>embB</i>	Y333H	EMB	
4247513	<i>embB</i>	Y334H	EMB	
4247552	<i>embB</i>	S347C	EMB	
4247553	<i>embB</i>	S347I	EMB	
4247553	<i>embB</i>	S347T	EMB	
4247573	<i>embB</i>	D354N	EMB	
4247574	<i>embB</i>	D354A	EMB	
4247579	<i>embB</i>	A356S	EMB	
4247580	<i>embB</i>	A356V	EMB	
4247582	<i>embB</i>	A357S	EMB	
4247583	<i>embB</i>	A357V	EMB	
4247586	<i>embB</i>	G358V	EMB	
4247588	<i>embB</i>	L359I	EMB	
4247591	<i>embB</i>	V360M	EMB	
4247609	<i>embB</i>	S366P	EMB	
4247613	<i>embB</i>	R367P	EMB	
4247615	<i>embB</i>	E368Q	EMB	
4247616	<i>embB</i>	E368A	EMB	
4247617	<i>embB</i>	E368D	EMB	
4247618	<i>embB</i>	V369L	EMB	
4247619	<i>embB</i>	V369A	EMB	
4247622	<i>embB</i>	L370R	EMB	
4247625	<i>embB</i>	P371R	EMB	
4247634	<i>embB</i>	G374V	EMB	
4247636	<i>embB</i>	P375A	EMB	
4247642	<i>embB</i>	V377M	EMB	
4247643	<i>embB</i>	V377G	EMB	
4247645	<i>embB</i>	E378K	EMB	

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

4247648	<i>embB</i>	A379T	EMB			
4247649	<i>embB</i>	A379D	EMB			
4247651	<i>embB</i>	S380R	EMB			
4247652	<i>embB</i>	S380N	EMB			
4247653	<i>embB</i>	S380R	EMB			
4247663	<i>embB</i>	Y384N	EMB			
4247676	<i>embB</i>	A388G	EMB			
4247696	<i>embB</i>	W395R	EMB			
4247698	<i>embB</i>	W395C	EMB			
4247702	<i>embB</i>	P397T	EMB			
4247703	<i>embB</i>	P397Q	EMB			
4247706	<i>embB</i>	F398Y	EMB			
4247708	<i>embB</i>	N399H	EMB			
4247708	<i>embB</i>	N399D	EMB			
4247713	<i>embB</i>	N400K	EMB			
4247714	<i>embB</i>	G401S	EMB			
4247717	<i>embB</i>	L402V	EMB			
4247723	<i>embB</i>	P404S	EMB			
4247728	<i>embB</i>	E405D	EMB			
4247729	<i>embB</i>	G406S	EMB			
4247729	<i>embB</i>	G406C	EMB			
4247730	<i>embB</i>	G406D	EMB			
4247730	<i>embB</i>	G406A	EMB			
4247748	<i>embB</i>	S412L	EMB			
4247782	<i>embB</i>	M423I	EMB			
4247802	<i>embB</i>	P430L	EMB			
4247804	<i>embB</i>	A431T	EMB			
4247817	<i>embB</i>	V435G	EMB			
4247820	<i>embB</i>	V436G	EMB			
4247822	<i>embB</i>	T437A	EMB			
4247850	<i>embB</i>	P446H	EMB			
4247856	<i>embB</i>	G448V	EMB			
4247863	<i>embB</i>	I450M	EMB			
4247867	<i>embB</i>	V452L	EMB			
4247873	<i>embB</i>	A454T	EMB			

D Mutationslisten

4247889	<i>embB</i>	G459A	EMB		
4247891	<i>embB</i>	R460C	EMB		
4247892	<i>embB</i>	R460L	EMB		
4247894	<i>embB</i>	P461S	EMB		
4247919	<i>embB</i>	R469P	EMB		
4247925	<i>embB</i>	R471P	EMB		
4247959	<i>embB</i>	M482I	EMB		
4248002	<i>embB</i>	Q497K	EMB		
4248003	<i>embB</i>	Q497P	EMB		
4248003	<i>embB</i>	Q497R	EMB		
4248025	<i>embB</i>	E504D	EMB		
4248033	<i>embB</i>	R507K	EMB		
4248488	<i>embB</i>	A659T	EMB		
4248747	<i>embB</i>	G745D	EMB		
4249071	<i>embB</i>	Q853P	EMB		
4249389	<i>embB</i>	D959A	EMB		
4249512	<i>embB</i>	M1000R	EMB		
4249518	<i>embB</i>	H1002R	EMB		
4249583	<i>embB</i>	D1024N	EMB		
4249757	<i>embB</i>	T1082A	EMB		
4326087	<i>ethA</i>	R463S	ETH		
4326093	<i>ethA</i>	Y461H	ETH		
4326150	<i>ethA</i>	AGC- 1324A	ETH		
4326183	<i>ethA</i>	AG1291A	ETH		
4326236	<i>ethA</i>	G413D	ETH		
4326267	<i>ethA</i>	C403G	ETH		
4326284	<i>ethA</i>	L397R	ETH		
4326300	<i>ethA</i>	T392A	ETH		
4326317	<i>ethA</i>	Y386C	ETH		
4326320	<i>ethA</i>	G385D	ETH		
4326333	<i>ethA</i>	A381P	ETH		
4326449	<i>ethA</i>	T342K	ETH		
4326461	<i>ethA</i>	I338S	ETH		
4326476	<i>ethA</i>	L333R	ETH		

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

4326488	<i>ethA</i>	S329L	ETH	
4326659	<i>ethA</i>	L272P	ETH	
4326669	<i>ethA</i>	Q269Stop	ETH	
4326705	<i>ethA</i>	GC769G	ETH	
4326738	<i>ethA</i>	Q246Stop	ETH	
4326770	<i>ethA</i>	TA704T	ETH	
4326780	<i>ethA</i>	T232A	ETH	
4326807	<i>ethA</i>	E223K	ETH	
4326917	<i>ethA</i>	T186K	ETH	
4326980	<i>ethA</i>	Q165P	ETH	
4327224	<i>ethA</i>	Y84D	ETH	
4327292	<i>ethA</i>	T61M	ETH	
4327301	<i>ethA</i>	D58A	ETH	
4327307	<i>ethA</i>	D56A	ETH	
4327322	<i>ethA</i>	P51L	ETH	
4327346	<i>ethA</i>	G43D	ETH	
4327347	<i>ethA</i>	G43C	ETH	
4327347	<i>ethA</i>	G43S	ETH	
4327363	<i>ethA</i>	CT111C	ETH	
4327409	<i>ethA</i>	H22P	ETH	
4327442	<i>ethA</i>	G11A	ETH	
4327472	<i>ethA</i>	M1R	ETH	
4327831	<i>ethR</i>	A95T	ETH	
4327876	<i>ethR</i>	F110L	ETH	
4327878	<i>ethR</i>	F110L	ETH	
4407604	<i>gidB</i>	A200E	SM	
4407790	<i>gidB</i>	A138V	SM	
4407824	<i>gidB</i>	Q127Stop	SM	
4407931	<i>gidB</i>	L91P	SM	
4407940	<i>gidB</i>	V88A	SM	
4407992	<i>gidB</i>	G71R	SM	
4408009	<i>gidB</i>	V65G	SM	
4408102	<i>gidB</i>	G34A	SM	
-	<i>embB</i>	G406P	EMB	
-	<i>embR</i>	L303E	EMB	

D Mutationslisten

-	<i>gyrA</i>	D94F	FQ		
-	<i>ahpC</i>	L3K	INH		
-	<i>katG</i>	G309F	INH		
-	<i>pncA</i>	F94P	PZA		
-	<i>rpoB</i>	S450Y	RMP		
1674484/ 1674485	<i>inhA</i>	I95P	ETH		
1674484/ 1674485	<i>inhA</i>	I95P	INH		
2154752/ 2154751/ 2154750	<i>katG</i>	E454R	INH		
2154887/ 2154886	<i>katG</i>	A409R	INH		
2155150/ 2155149	<i>katG</i>	W321S	INH		
2155159/ 2155158	<i>katG</i>	E318V	INH		
2155168/ 2155167	<i>katG</i>	S315T	INH		
2155169/ 2155168/2155167	<i>katG</i>	S315L	INH		
2155273/ 2155272	<i>katG</i>	P280H	INH		
2155556/ 2155555	<i>katG</i>	G186H	INH		
2155790/ 2155789/ 2155788	<i>katG</i>	H108E	INH		
2288728/ 2288727	<i>pncA</i>	L172A	PZA		
2288962/ 2288961	<i>pncA</i>	F94P	PZA		
2289168/ 2289167	<i>pncA</i>	A25E	PZA		

D.1 Eingeschlossenen resistenzvermittelnde Mutationen pro untersuchtem Programm

2289174/ 2289173	<i>pncA</i>	G23V	PZA			
2289213/ 2289212	<i>pncA</i>	Q10P	PZA			
2289217/ 2289216	<i>pncA</i>	V9S	PZA			
761083/ 761084	<i>rpoB</i>	G426D	RMP			
761085/ 761086	<i>rpoB</i>	T427H	RMP			
761088/ 761089/ 761090	<i>rpoB</i>	S428Q	RMP			
761100/ 761101/ 761102	<i>rpoB</i>	Q432N	RMP			
761109/ 761110	<i>rpoB</i>	D435F	RMP			
761110/ 761111	<i>rpoB</i>	D435V	RMP			
761127/ 761128	<i>rpoB</i>	S441Q	RMP			
761139/ 761140	<i>rpoB</i>	H445F	RMP			
761139/ 761140	<i>rpoB</i>	H445G	RMP			
761140/ 761141	<i>rpoB</i>	H445L	RMP			
761146/ 761147	<i>rpoB</i>	R447P	RMP			
761154/ 761155	<i>rpoB</i>	S450Q	RMP			
761155/ 761156	<i>rpoB</i>	S450F	RMP			

D Mutationslisten

761155/ 761156	<i>rpoB</i>	S450Y	RMP	■	■
761161/ 761162	<i>rpoB</i>	L452P	RMP	■	■

E Berechnung von Sensitivität und Spezifität

E.1 Vier-Felder Tafeln zur Berechnung der Sensitivitäten und Spezifitäten basierend auf den Sanger-Ergebnissen)

E.1.1 Isoniazid Resistenzbestimmung

CASTB Isoniazid Vorhersage

```
> CASTBINH <- as.table(matrix(c(24,0,5,62), nrow = 2, byrow = TRUE))
> rval <- epi.tests(CASTBINH, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
CASTB +	24	0	24
CASTB -	5	62	67
Total	29	62	91

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,26 (0,18; 0,37)
Wahre Prävalenz	0,32 (0,22; 0,42)
Sensitivität	0,83 (0,64; 0,94)
Spezifität	1,00 (0,91; 1,00)
Positiver prädiktiver Wert	1,00 (0,80; 1,00)
Negativer prädiktiver Wert	0,93 (0,83; 0,98)

```
Positiver Wahrscheinlichkeitsquotient    Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient    0,17 (0,08; 0,38)
-----
```

KvarQ Isoniazid Vorhersage

```
> KvarQINH <- as.table(matrix(c(24,0,5,62), nrow = 2, byrow = TRUE))
> rval <- epi.tests(KvarQINH, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
KvarQ +	24	0	24
KvarQ -	5	62	67
Total	29	62	91

Punktschätzer und 95% KIs:

```
-----
Beobachtete Prävalenz                    0,26 (0,18; 0,37)
Wahre Prävalenz                          0,32 (0,22; 0,42)
Sensitivität                             0,83 (0,64; 0,94)
Spezifität                                1,00 (0,91; 1,00)
Positiver prädiktiver Wert                1,00 (0,80; 1,00)
Negativer prädiktiver Wert                0,93 (0,83; 0,98)
Positiver Wahrscheinlichkeitsquotient    Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient    0,17 (0,08; 0,38)
-----
```

Mykrobe Predictor TB Isoniazid Vorhersage

```
> dat <- as.table(matrix(c(25,0,4,62), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
Myk Prd +	25	0	25
Myk Prd -	4	62	66

*E.1 Vier-Felder Tafeln zur Berechnung der Sensitivitäten und Spezifitäten
basierend auf den Sanger-Ergebnissen)*

Total	29	62	91
-------	----	----	----

Punktschätzer und 95% KIs:

```
-----
Beobachtete Prävalenz           0,27 (0,19; 0,38)
Wahre Prävalenz                 0,32 (0,22; 0,42)
Sensitivität                    0,86 (0,68; 0,96)
Spezifität                      1,00 (0,91; 1,00)
Positiver prädiktiver Wert      1,00 (0,80; 1,00)
Negativer prädiktiver Wert      0,94 (0,85; 0,98)
Positiver Wahrscheinlichkeitsquotient  Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient  0,14 (0,06; 0,34)
-----
```

PhyResSE Isoniazid Vorhersage

```
> dat <- as.table(matrix(c(29,0,0,62), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
PhyResSE +	29	0	29
PhyResSE -	0	62	62
Total	29	62	91

Punktschätzer und 95% KIs:

```
-----
Beobachtete Prävalenz           0,32 (0,22; 0,42)
Wahre Prävalenz                 0,32 (0,22; 0,42)
Sensitivität                    1,00 (0,83; 1,00)
Spezifität                      1,00 (0,91; 1,00)
Positiver prädiktiver Wert      1,00 (0,83; 1,00)
Negativer prädiktiver Wert      1,00 (0,91; 1,00)
Positiver Wahrscheinlichkeitsquotient  Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient  0,00 (0,00; NaN)
-----
```

TBProfiler Isoniazid Vorhersage

```
> dat <- as.table(matrix(c(27,9,2,53), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
TBProfiler +	27	9	36
TBProfiler -	2	53	55
Total	29	62	91

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,40 (0,29; 0,50)
Wahre Prävalenz	0,32 (0,22; 0,42)
Sensitivität	0,93 (0,77; 0,99)
Spezifität	0,85 (0,74; 0,93)
Positiver prädiktiver Wert	0,75 (0,58; 0,88)
Negativer prädiktiver Wert	0,96 (0,87; 1,00)
Positiver Wahrscheinlichkeitsquotient	6,41 (3,48; 11,83)
Negativer Wahrscheinlichkeitsquotient	0,08 (0,02; 0,32)

E.1.2 Rifampicin Resistenzbestimmung

CASTB Rifampicin Vorhersage

```
> dat <- as.table(matrix(c(17,0,1,73), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
CASTB +	17	0	17
CASTB -	1	73	74
Total	18	73	91

*E.1 Vier-Felder Tafeln zur Berechnung der Sensitivitäten und Spezifitäten
basierend auf den Sanger-Ergebnissen)*

Punktschätzer und 95% KIs:

```
-----
```

Beobachtete Prävalenz	0,19 (0,11; 0,28)
Wahre Prävalenz	0,20 (0,12; 0,29)
Sensitivität	0,94 (0,73; 1,00)
Spezifität	1,00 (0,93; 1,00)
Positiver prädiktiver Wert	1,00 (0,73; 1,00)
Negativer prädiktiver Wert	0,99 (0,93; 1,00)
Positiver Wahrscheinlichkeitsquotient	Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient	0,06 (0,01; 0,37)

```
-----
```

KvarQ Rifampicin Vorhersage

```
> dat <- as.table(matrix(c(17,0,1,73), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
KvarQ +	17	0	17
KvarQ -	1	73	74
Total	18	73	91

Punktschätzer und 95% KIs:

```
-----
```

Beobachtete Prävalenz	0,19 (0,11; 0,28)
Wahre Prävalenz	0,20 (0,12; 0,29)
Sensitivität	0,94 (0,73; 1,00)
Spezifität	1,00 (0,93; 1,00)
Positiver prädiktiver Wert	1,00 (0,73; 1,00)
Negativer prädiktiver Wert	0,99 (0,93; 1,00)
Positiver Wahrscheinlichkeitsquotient	Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient	0,06 (0,01; 0,37)

```
-----
```

Mykrobe Predictor TB Rifampicin Vorhersage

```
> dat <- as.table(matrix(c(18,1,0,72), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
Myk Prd +	18	1	19
Myk Prd -	0	72	72
Total	18	73	91

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,21 (0,13; 0,31)
Wahre Prävalenz	0,20 (0,12; 0,29)
Sensitivität	1,00 (0,74; 1,00)
Spezifität	0,99 (0,93; 1,00)
Positiver prädiktiver Wert	0,95 (0,74; 1,00)
Negativer prädiktiver Wert	1,00 (0,93; 1,00)
Positiver Wahrscheinlichkeitsquotient	73,00 (10,42; 511,30)
Negativer Wahrscheinlichkeitsquotient	0,00 (0,00; NaN)

PhyResSE Rifampicin Vorhersage

```
> dat <- as.table(matrix(c(18,1,0,72), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
PhyResSE +	18	1	19
PhyResSE -	0	72	72
Total	18	73	91

Punktschätzer und 95% KIs:

*E.1 Vier-Felder Tafeln zur Berechnung der Sensitivitäten und Spezifitäten
basierend auf den Sanger-Ergebnissen)*

Beobachtete Prävalenz	0,21 (0,13; 0,31)
Wahre Prävalenz	0,20 (0,12; 0,29)
Sensitivität	1,00 (0,74; 1,00)
Spezifität	0,99 (0,93; 1,00)
Positiver prädiktiver Wert	0,95 (0,74; 1,00)
Negativer prädiktiver Wert	1,00 (0,93; 1,00)
Positiver Wahrscheinlichkeitsquotient	73,00 (10,42; 511,30)
Negativer Wahrscheinlichkeitsquotient	0,00 (0,00; NaN)

TBProfiler Rifampicin Vorhersage

```
> dat <- as.table(matrix(c(18,1,0,72), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
TBProfiler +	18	1	19
TBProfiler -	0	72	72
Total	18	73	91

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,21 (0,13; 0,31)
Wahre Prävalenz	0,20 (0,12; 0,29)
Sensitivität	1,00 (0,74; 1,00)
Spezifität	0,99 (0,93; 1,00)
Positiver prädiktiver Wert	0,95 (0,74; 1,00)
Negativer prädiktiver Wert	1,00 (0,93; 1,00)
Positiver Wahrscheinlichkeitsquotient	73,00 (10,42; 511,30)
Negativer Wahrscheinlichkeitsquotient	0,00 (0,00; NaN)

E.1.3 Ethambutol Resistenzbestimmung

CASTB Ethambutol Vorhersage

```
> dat <- as.table(matrix(c(8,0,7,76), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
CASTB +	8	0	8
CASTB -	7	76	83
Total	15	76	91

Punktschätzer und 95% KIs:

```
-----
Beobachtete Prävalenz          0,09 (0,04; 0,17)
Wahre Prävalenz                0,16 (0,10; 0,26)
Sensitivität                  0,53 (0,27; 0,79)
Spezifität                    1,00 (0,93; 1,00)
Positiver prädiktiver Wert    1,00 (0,52; 1,00)
Negativer prädiktiver Wert    0,92 (0,83; 0,97)
Positiver Wahrscheinlichkeitsquotient  Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient  0,47 (0,27; 0,80)
-----
```

KvarQ Ethambutol Vorhersage

```
> dat <- as.table(matrix(c(8,0,7,76), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
KvarQ +	8	0	8
KvarQ -	7	76	83
Total	15	76	91

*E.1 Vier-Felder Tafeln zur Berechnung der Sensitivitäten und Spezifitäten
basierend auf den Sanger-Ergebnissen)*

Punktschätzer und 95% KIs:

```
-----
```

Beobachtete Prävalenz	0,09 (0,04; 0,17)
Wahre Prävalenz	0,16 (0,10; 0,26)
Sensitivität	0,53 (0,27; 0,79)
Spezifität	1,00 (0,93; 1,00)
Positiver prädiktiver Wert	1,00 (0,52; 1,00)
Negativer prädiktiver Wert	0,92 (0,83; 0,97)
Positiver Wahrscheinlichkeitsquotient	Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient	0,47 (0,27; 0,80)

```
-----
```

Mykrobe Predictor TB Ethambutol Vorhersage

```
> dat <- as.table(matrix(c(7,1,8,75), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
Myk Prd +	7	1	8
Myk Prd -	8	75	83
Total	15	76	91

Punktschätzer und 95% KIs:

```
-----
```

Beobachtete Prävalenz	0,09 (0,04; 0,17)
Wahre Prävalenz	0,16 (0,10; 0,26)
Sensitivität	0,47 (0,21; 0,73)
Spezifität	0,99 (0,93; 1,00)
Positiver prädiktiver Wert	0,88 (0,47; 1,00)
Negativer prädiktiver Wert	0,90 (0,82; 0,96)
Positiver Wahrscheinlichkeitsquotient	35,47 (4,70; 267,57)
Negativer Wahrscheinlichkeitsquotient	0,54 (0,34; 0,87)

```
-----
```

PhyResSE Ethambutol Vorhersage

```
> dat <- as.table(matrix(c(14,0,1,76), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
PhyResSE +	14	0	14
PhyResSE -	1	76	77
Total	15	76	91

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,15 (0,09; 0,24)
Wahre Prävalenz	0,16 (0,10; 0,26)
Sensitivität	0,93 (0,68; 1,00)
Spezifität	1,00 (0,93; 1,00)
Positiver prädiktiver Wert	1,00 (0,68; 1,00)
Negativer prädiktiver Wert	0,99 (0,93; 1,00)
Positiver Wahrscheinlichkeitsquotient	Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient	0,07 (0,01; 0,44)

TBProfiler Ethambutol Vorhersage

```
> dat <- as.table(matrix(c(14,1,1,75), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
TBProfiler +	14	1	15
TBProfiler -	1	75	76
Total	15	76	91

Punktschätzer und 95% KIs:

*E.1 Vier-Felder Tafeln zur Berechnung der Sensitivitäten und Spezifitäten
basierend auf den Sanger-Ergebnissen)*

Beobachtete Prävalenz	0,16 (0,10; 0,26)
Wahre Prävalenz	0,16 (0,10; 0,26)
Sensitivität	0,93 (0,68; 1,00)
Spezifität	0,99 (0,93; 1,00)
Positiver prädiktiver Wert	0,93 (0,68; 1,00)
Negativer prädiktiver Wert	0,99 (0,93; 1,00)
Positiver Wahrscheinlichkeitsquotient	70,93 (10,07; 499,43)
Negativer Wahrscheinlichkeitsquotient	0,07 (0,01; 0,45)

E.1.4 Streptomycin Resistenzbestimmung

CASTB Streptomycin Vorhersage

```
> dat <- as.table(matrix(c(11,0,26,54), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
CASTB +	11	0	11
CASTB -	26	54	80
Total	37	54	91

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,12 (0,06; 0,21)
Wahre Prävalenz	0,41 (0,30; 0,51)
Sensitivität	0,30 (0,16; 0,47)
Spezifität	1,00 (0,90; 1,00)
Positiver prädiktiver Wert	1,00 (0,62; 1,00)
Negativer prädiktiver Wert	0,68 (0,56; 0,78)
Positiver Wahrscheinlichkeitsquotient	Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient	0,70 (0,57; 0,87)

KvarQ Streptomycin Vorhersage

```
> dat <- as.table(matrix(c(21,0,16,54), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
KvarQ +	21	0	21
KvarQ -	16	54	70
Total	37	54	91

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,23 (0,15; 0,33)
Wahre Prävalenz	0,41 (0,30; 0,51)
Sensitivität	0,57 (0,39; 0,73)
Spezifität	1,00 (0,90; 1,00)
Positiver prädiktiver Wert	1,00 (0,77; 1,00)
Negativer prädiktiver Wert	0,77 (0,66; 0,86)
Positiver Wahrscheinlichkeitsquotient	Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient	0,43 (0,30; 0,63)

Mykrobe Predictor TB Streptomycin Vorhersage

```
> dat <- as.table(matrix(c(21,0,16,54), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
Myk Prd +	21	0	21
Myk Prd -	16	54	70
Total	37	54	91

Punktschätzer und 95% KIs:

*E.1 Vier-Felder Tafeln zur Berechnung der Sensitivitäten und Spezifitäten
basierend auf den Sanger-Ergebnissen)*

Beobachtete Prävalenz	0,23 (0,15; 0,33)
Wahre Prävalenz	0,41 (0,30; 0,51)
Sensitivität	0,57 (0,39; 0,73)
Spezifität	1,00 (0,90; 1,00)
Positiver prädiktiver Wert	1,00 (0,77; 1,00)
Negativer prädiktiver Wert	0,77 (0,66; 0,86)
Positiver Wahrscheinlichkeitsquotient	Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient	0,43 (0,30; 0,63)

PhyResSE Streptomycin Vorhersage

```
> dat <- as.table(matrix(c(32,0,5,54), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
PhyResSE +	32	0	32
PhyResSE -	5	54	59
Total	37	54	91

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,35 (0,25; 0,46)
Wahre Prävalenz	0,41 (0,30; 0,51)
Sensitivität	0,86 (0,71; 0,95)
Spezifität	1,00 (0,90; 1,00)
Positiver prädiktiver Wert	1,00 (0,84; 1,00)
Negativer prädiktiver Wert	0,92 (0,81; 0,97)
Positiver Wahrscheinlichkeitsquotient	Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient	0,14 (0,06; 0,31)

TBProfiler Streptomycin Vorhersage

```
> dat <- as.table(matrix(c(21,0,16,54), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
TBProfiler +	21	0	21
TBProfiler -	16	54	70
Total	37	54	91

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,23 (0,15; 0,33)
Wahre Prävalenz	0,41 (0,30; 0,51)
Sensitivität	0,57 (0,39; 0,73)
Spezifität	1,00 (0,90; 1,00)
Positiver prädiktiver Wert	1,00 (0,77; 1,00)
Negativer prädiktiver Wert	0,77 (0,66; 0,86)
Positiver Wahrscheinlichkeitsquotient	Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient	0,43 (0,30; 0,63)

E.1.5 Pyrazinamid Resistenzbestimmung

CASTB Pyrazinamid Vorhersage

```
> dat <- as.table(matrix(c(5,0,7,79), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
CASTB +	5	0	5
CASTB -	7	79	86
Total	12	79	91

*E.1 Vier-Felder Tafeln zur Berechnung der Sensitivitäten und Spezifitäten
basierend auf den Sanger-Ergebnissen)*

Punktschätzer und 95% KIs:

```
-----
```

Beobachtete Prävalenz	0,05 (0,02; 0,12)
Wahre Prävalenz	0,13 (0,07; 0,22)
Sensitivität	0,42 (0,15; 0,72)
Spezifität	1,00 (0,93; 1,00)
Positiver prädiktiver Wert	1,00 (0,36; 1,00)
Negativer prädiktiver Wert	0,92 (0,84; 0,97)
Positiver Wahrscheinlichkeitsquotient	Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient	0,58 (0,36; 0,94)

```
-----
```

KvarQ Pyrazinamid Vorhersage

```
> dat <- as.table(matrix(c(5,0,7,79), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
Test +	5	0	5
Test -	7	79	86
Total	12	79	91

Punktschätzer und 95% KIs:

```
-----
```

Beobachtete Prävalenz	0,05 (0,02; 0,12)
Wahre Prävalenz	0,13 (0,07; 0,22)
Sensitivität	0,42 (0,15; 0,72)
Spezifität	1,00 (0,93; 1,00)
Positiver prädiktiver Wert	1,00 (0,36; 1,00)
Negativer prädiktiver Wert	0,92 (0,84; 0,97)
Positiver Wahrscheinlichkeitsquotient	Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient	0,58 (0,36; 0,94)

```
-----
```

Mykrobe Predictor TB Pyrazinamid Vorhersage

Nicht verfügbar.

PhyResSE Pyrazinamid Vorhersage

```
> dat <- as.table(matrix(c(9,0,3,79), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
Myk Prd +	9	0	9
Myk Prd -	3	79	82
Total	12	79	91

Punktschätzer und 95% KIs:

```
-----
```

Beobachtete Prävalenz	0,10 (0,05; 0,18)
Wahre Prävalenz	0,13 (0,07; 0,22)
Sensitivität	0,75 (0,43; 0,95)
Spezifität	1,00 (0,93; 1,00)
Positiver prädiktiver Wert	1,00 (0,55; 1,00)
Negativer prädiktiver Wert	0,96 (0,90; 0,99)
Positiver Wahrscheinlichkeitsquotient	Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient	0,25 (0,09; 0,67)

```
-----
```

TBProfiler Pyrazinamid Vorhersage

```
> dat <- as.table(matrix(c(7,0,5,79), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	Sanger +	Sanger -	Total
TBProfiler +	7	0	7
TBProfiler -	5	79	84

*E.2 Vier-Felder Tafeln zur Berechnung der Sensitivitäten und Spezifitäten
basierend auf der phänotypischen Resistenztestung)*

Total	12	79	91
-------	----	----	----

Punktschätzer und 95% KIs:

```
-----
Beobachtete Prävalenz           0,08 (0,03; 0,15)
Wahre Prävalenz                 0,13 (0,07; 0,22)
Sensitivität                   0,58 (0,28; 0,85)
Spezifität                     1,00 (0,93; 1,00)
Positiver prädiktiver Wert     1,00 (0,47; 1,00)
Negativer prädiktiver Wert     0,94 (0,87; 0,98)
Positiver Wahrscheinlichkeitsquotient  Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient  0,42 (0,21; 0,81)
-----
```

E.2 Vier-Felder Tafeln zur Berechnung der Sensitivitäten und Spezifitäten basierend auf der phänotypischen Resistenztestung)

E.2.1 Isoniazid Resistenzbestimmung

CASTB Isoniazid Vorhersage

```
> CASTBINH <- as.table(matrix(c(23,1,6,61), nrow = 2, byrow = TRUE))
> rval <- epi.tests(CASTBINH, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
CASTB +	24	1	25
CASTB -	5	61	66
Total	29	62	91

Punktschätzer und 95% KIs:

```
-----
Beobachtete Prävalenz           0,27 (0,19; 0,38)
```

E Berechnung von Sensitivität und Spezifität

Wahre Prävalenz	0,32 (0,22; 0,42)
Sensitivität	0,83 (0,64; 0,94)
Spezifität	0,98 (0,91; 1,00)
Positiver prädiktiver Wert	0,96 (0,80; 1,00)
Negativer prädiktiver Wert	0,92 (0,83; 0,97)
Positiver Wahrscheinlichkeitsquotient	51,31 (7,29; 361,07)
Negativer Wahrscheinlichkeitsquotient	0,18 (0,08; 0,39)

KvarQ Isoniazid Vorhersage

```
> KvarQINH <- as.table(matrix(c(23,1,6,61), nrow = 2, byrow = TRUE))
> rval <- epi.tests(KvarQINH, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
KvarQ +	23	1	24
KvarQ -	6	61	67
Total	29	62	91

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,26 (0,18; 0,37)
Wahre Prävalenz	0,32 (0,22; 0,42)
Sensitivität	0,79 (0,60; 0,92)
Spezifität	0,98 (0,91; 1,00)
Positiver prädiktiver Wert	0,96 (0,79; 1,00)
Negativer prädiktiver Wert	0,91 (0,82; 0,97)
Positiver Wahrscheinlichkeitsquotient	49,17 (6,98; 346,64)
Negativer Wahrscheinlichkeitsquotient	0,21 (0,10; 0,43)

Mykrobe Predictor TB Isoniazid Vorhersage

```
> dat <- as.table(matrix(c(23,3,6,59), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
Myk Prd +	23	3	26
Myk Prd -	6	59	65
Total	29	62	91

Punktschätzer und 95% KIs:

```
-----
```

Beobachtete Prävalenz	0,29 (0,20; 0,39)
Wahre Prävalenz	0,32 (0,22; 0,42)
Sensitivität	0,79 (0,60; 0,92)
Spezifität	0,95 (0,87; 0,99)
Positiver prädiktiver Wert	0,88 (0,70; 0,98)
Negativer prädiktiver Wert	0,91 (0,81; 0,97)
Positiver Wahrscheinlichkeitsquotient	16,39 (5,35; 50,21)
Negativer Wahrscheinlichkeitsquotient	0,22 (0,11; 0,44)

```
-----
```

PhyResSE Isoniazid Vorhersage

```
> dat <- as.table(matrix(c(27,2,2,60), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
PhyResSE +	27	2	29
PhyResSE -	2	60	62
Total	29	62	91

Punktschätzer und 95% KIs:

```
-----
```

E Berechnung von Sensitivität und Spezifität

Beobachtete Prävalenz	0,32 (0,22; 0,42)
Wahre Prävalenz	0,32 (0,22; 0,42)
Sensitivität	0,93 (0,77; 0,99)
Spezifität	0,97 (0,89; 1,00)
Positiver prädiktiver Wert	0,93 (0,77; 0,99)
Negativer prädiktiver Wert	0,97 (0,89; 1,00)
Positiver Wahrscheinlichkeitsquotient	28,86 (7,36; 113,24)
Negativer Wahrscheinlichkeitsquotient	0,07 (0,02; 0,27)

TBProfiler Isoniazid Vorhersage

```
> dat <- as.table(matrix(c(26,10,3,52), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
TBProfiler +	26	10	36
TBProfiler -	3	52	55
Total	29	62	91

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,40 (0,29; 0,50)
Wahre Prävalenz	0,32 (0,22; 0,42)
Sensitivität	0,90 (0,73; 0,98)
Spezifität	0,84 (0,72; 0,92)
Positiver prädiktiver Wert	0,72 (0,55; 0,86)
Negativer prädiktiver Wert	0,95 (0,85; 0,99)
Positiver Wahrscheinlichkeitsquotient	5,56 (3,11; 9,94)
Negativer Wahrscheinlichkeitsquotient	0,12 (0,04; 0,36)

E.2.2 Rifampicin Resistenzbestimmung

CASTB Rifampicin Vorhersage

```
> dat <- as.table(matrix(c(14,3,0,74), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
CASTB +	14	3	17
CASTB -	0	74	74
Total	14	77	91

Punktschätzer und 95% KIs:

```
-----
```

Beobachtete Prävalenz	0,19 (0,11; 0,28)
Wahre Prävalenz	0,15 (0,09; 0,24)
Sensitivität	1,00 (0,68; 1,00)
Spezifität	0,96 (0,89; 0,99)
Positiver prädiktiver Wert	0,82 (0,57; 0,96)
Negativer prädiktiver Wert	1,00 (0,93; 1,00)
Positiver Wahrscheinlichkeitsquotient	25,67 (8,46; 77,83)
Negativer Wahrscheinlichkeitsquotient	0,00 (0,00; NaN)

```
-----
```

KvarQ Rifampicin Vorhersage

```
> dat <- as.table(matrix(c(13,4,1,73), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
KvarQ +	13	4	17
KvarQ -	1	73	74
Total	14	77	91

Punktschätzer und 95% KIs:

```
-----
```

Beobachtete Prävalenz	0,19 (0,11; 0,28)
Wahre Prävalenz	0,15 (0,09; 0,24)
Sensitivität	0,93 (0,66; 1,00)
Spezifität	0,95 (0,87; 0,99)
Positiver prädiktiver Wert	0,76 (0,50; 0,93)
Negativer prädiktiver Wert	0,99 (0,93; 1,00)
Positiver Wahrscheinlichkeitsquotient	17,87 (6,81; 46,93)
Negativer Wahrscheinlichkeitsquotient	0,08 (0,01; 0,50)

```
-----
```

Mykrobe Predictor TB Rifampicin Vorhersage

```
> dat <- as.table(matrix(c(14,5,0,72), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
Myk Prd +	14	5	19
Myk Prd -	0	72	72
Total	14	77	91

Punktschätzer und 95% KIs:

```
-----
```

Beobachtete Prävalenz	0,21 (0,13; 0,31)
Wahre Prävalenz	0,15 (0,09; 0,24)
Sensitivität	1,00 (0,68; 1,00)
Spezifität	0,94 (0,85; 0,98)
Positiver prädiktiver Wert	0,74 (0,49; 0,91)
Negativer prädiktiver Wert	1,00 (0,93; 1,00)
Positiver Wahrscheinlichkeitsquotient	15,40 (6,60; 35,94)
Negativer Wahrscheinlichkeitsquotient	0,00 (0,00; NaN)

```
-----
```

PhyResSE Rifampicin Vorhersage

```
> dat <- as.table(matrix(c(14,5,0,72), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
PhyResSE +	14	5	19
PhyResSE -	0	72	72
Total	14	77	91

Punktschätzer und 95% KIs:

```
-----
```

Beobachtete Prävalenz	0,21 (0,13; 0,31)
Wahre Prävalenz	0,15 (0,09; 0,24)
Sensitivität	1,00 (0,68; 1,00)
Spezifität	0,94 (0,85; 0,98)
Positiver prädiktiver Wert	0,74 (0,49; 0,91)
Negativer prädiktiver Wert	1,00 (0,93; 1,00)
Positiver Wahrscheinlichkeitsquotient	15,40 (6,60; 35,94)
Negativer Wahrscheinlichkeitsquotient	0,00 (0,00; NaN)

```
-----
```

TBProfiler Rifampicin Vorhersage

```
> dat <- as.table(matrix(c(14,5,0,72), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
TBProfiler +	14	5	19
TBProfiler -	0	72	72
Total	14	77	91

Punktschätzer und 95% KIs:

```
-----
```

Beobachtete Prävalenz	0,21 (0,13; 0,31)
Wahre Prävalenz	0,15 (0,09; 0,24)
Sensitivität	1,00 (0,68; 1,00)
Spezifität	0,94 (0,85; 0,98)
Positiver prädiktiver Wert	0,74 (0,49; 0,91)
Negativer prädiktiver Wert	1,00 (0,93; 1,00)
Positiver Wahrscheinlichkeitsquotient	15,40 (6,60; 35,94)
Negativer Wahrscheinlichkeitsquotient	0,00 (0,00; NaN)

E.2.3 Ethambutol Resistenzbestimmung

CASTB Ehtambutol Vorhersage

```
> dat <- as.table(matrix(c(8,0,6,77), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
CASTB +	8	0	8
CASTB -	6	77	83
Total	14	77	91

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,09 (0,04; 0,17)
Wahre Prävalenz	0,15 (0,09; 0,24)
Sensitivität	0,57 (0,29; 0,82)
Spezifität	1,00 (0,93; 1,00)
Positiver prädiktiver Wert	1,00 (0,52; 1,00)
Negativer prädiktiver Wert	0,93 (0,85; 0,97)
Positiver Wahrscheinlichkeitsquotient	Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient	0,43 (0,23; 0,78)

KvarQ Ethambutol Vorhersage

```
> dat <- as.table(matrix(c(7,1,7,76), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
KvarQ +	7	1	8
KvarQ -	7	76	83
Total	14	77	91

Punktschätzer und 95% KIs:

```
-----
```

Beobachtete Prävalenz	0,09 (0,04; 0,17)
Wahre Prävalenz	0,15 (0,09; 0,24)
Sensitivität	0,50 (0,23; 0,77)
Spezifität	0,99 (0,93; 1,00)
Positiver prädiktiver Wert	0,88 (0,47; 1,00)
Negativer prädiktiver Wert	0,92 (0,83; 0,97)
Positiver Wahrscheinlichkeitsquotient	38,50 (5,13; 289,19)
Negativer Wahrscheinlichkeitsquotient	0,51 (0,30; 0,86)

```
-----
```

Mykrobe Predictor TB Ethambutol Vorhersage

```
> dat <- as.table(matrix(c(7,1,7,76), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
Myk Prd +	7	1	8
Myk Prd -	7	76	83
Total	14	77	91

Punktschätzer und 95% KIs:

```
-----
```

E Berechnung von Sensitivität und Spezifität

Beobachtete Prävalenz	0,09 (0,04; 0,17)
Wahre Prävalenz	0,15 (0,09; 0,24)
Sensitivität	0,50 (0,23; 0,77)
Spezifität	0,99 (0,93; 1,00)
Positiver prädiktiver Wert	0,88 (0,47; 1,00)
Negativer prädiktiver Wert	0,92 (0,83; 0,97)
Positiver Wahrscheinlichkeitsquotient	38,50 (5,13; 289,19)
Negativer Wahrscheinlichkeitsquotient	0,51 (0,30; 0,86)

PhyResSE Ethambutol Vorhersage

```
> dat <- as.table(matrix(c(12,2,2,75), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
PhyResSE +	12	2	14
PhyResSE -	2	75	77
Total	14	77	91

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,15 (0,09; 0,24)
Wahre Prävalenz	0,15 (0,09; 0,24)
Sensitivität	0,86 (0,57; 0,98)
Spezifität	0,97 (0,91; 1,00)
Positiver prädiktiver Wert	0,86 (0,57; 0,98)
Negativer prädiktiver Wert	0,97 (0,91; 1,00)
Positiver Wahrscheinlichkeitsquotient	33,00 (8,27; 131,75)
Negativer Wahrscheinlichkeitsquotient	0,15 (0,04; 0,53)

TBProfiler Ethambutol Vorhersage

```
> dat <- as.table(matrix(c(12,3,2,74), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
TBProfiler +	12	3	15
TBProfiler -	2	74	76
Total	14	77	91

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,16 (0,10; 0,26)
Wahre Prävalenz	0,15 (0,09; 0,24)
Sensitivität	0,86 (0,57; 0,98)
Spezifität	0,96 (0,89; 0,99)
Positiver prädiktiver Wert	0,80 (0,52; 0,96)
Negativer prädiktiver Wert	0,97 (0,91; 1,00)
Positiver Wahrscheinlichkeitsquotient	22,00 (7,11; 68,09)
Negativer Wahrscheinlichkeitsquotient	0,15 (0,04; 0,54)

E.2.4 Streptomycin Resistenzbestimmung

CASTB Streptomycin Vorhersage

```
> dat <- as.table(matrix(c(11,0,26,54), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
CASTB +	11	0	11
CASTB -	26	54	80
Total	37	54	91

Punktschätzer und 95% KIs:

```
-----
```

Beobachtete Prävalenz	0,12 (0,06; 0,21)
Wahre Prävalenz	0,41 (0,30; 0,51)
Sensitivität	0,30 (0,16; 0,47)
Spezifität	1,00 (0,90; 1,00)
Positiver prädiktiver Wert	1,00 (0,62; 1,00)
Negativer prädiktiver Wert	0,68 (0,56; 0,78)
Positiver Wahrscheinlichkeitsquotient	Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient	0,70 (0,57; 0,87)

```
-----
```

KvarQ Streptomycin Vorhersage

```
> dat <- as.table(matrix(c(21,0,16,54), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
KvarQ +	21	0	21
KvarQ -	16	54	70
Total	37	54	91

Punktschätzer und 95% KIs:

```
-----
```

Beobachtete Prävalenz	0,23 (0,15; 0,33)
Wahre Prävalenz	0,41 (0,30; 0,51)
Sensitivität	0,57 (0,39; 0,73)
Spezifität	1,00 (0,90; 1,00)
Positiver prädiktiver Wert	1,00 (0,77; 1,00)
Negativer prädiktiver Wert	0,77 (0,66; 0,86)
Positiver Wahrscheinlichkeitsquotient	Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient	0,43 (0,30; 0,63)

```
-----
```

Mykrobe Predictor TB Streptomycin Vorhersage

```
> dat <- as.table(matrix(c(21,0,16,54), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
Myk Prd +	21	0	21
Myk Prd -	16	54	70
Total	37	54	91

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,23 (0,15; 0,33)
Wahre Prävalenz	0,41 (0,30; 0,51)
Sensitivität	0,57 (0,39; 0,73)
Spezifität	1,00 (0,90; 1,00)
Positiver prädiktiver Wert	1,00 (0,77; 1,00)
Negativer prädiktiver Wert	0,77 (0,66; 0,86)
Positiver Wahrscheinlichkeitsquotient	Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient	0,43 (0,30; 0,63)

PhyResSE Streptomycin Vorhersage

```
> dat <- as.table(matrix(c(31,1,6,53), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
PhyResSE +	31	1	32
PhyResSE -	6	53	59
Total	37	54	91

Punktschätzer und 95% KIs:

E Berechnung von Sensitivität und Spezifität

Beobachtete Prävalenz	0,35 (0,25; 0,46)
Wahre Prävalenz	0,41 (0,30; 0,51)
Sensitivität	0,84 (0,68; 0,94)
Spezifität	0,98 (0,90; 1,00)
Positiver prädiktiver Wert	0,97 (0,84; 1,00)
Negativer prädiktiver Wert	0,90 (0,79; 0,96)
Positiver Wahrscheinlichkeitsquotient	45,24 (6,46; 317,02)
Negativer Wahrscheinlichkeitsquotient	0,17 (0,08; 0,34)

TBProfiler Streptomycin Vorhersage

```
> dat <- as.table(matrix(c(21,0,16,54), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
TBProfiler +	21	0	21
TBProfiler -	16	54	70
Total	37	54	91

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,23 (0,15; 0,33)
Wahre Prävalenz	0,41 (0,30; 0,51)
Sensitivität	0,57 (0,39; 0,73)
Spezifität	1,00 (0,90; 1,00)
Positiver prädiktiver Wert	1,00 (0,77; 1,00)
Negativer prädiktiver Wert	0,77 (0,66; 0,86)
Positiver Wahrscheinlichkeitsquotient	Inf (NaN; Inf)
Negativer Wahrscheinlichkeitsquotient	0,43 (0,30; 0,63)

E.2.5 Pyrazinamid Resistenzbestimmung

CASTB Pyrazinamid Vorhersage

```
> dat <- as.table(matrix(c(4,1,5,81), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
CASTB +	4	1	5
CASTB -	5	81	86
Total	9	82	91

Punktschätzer und 95% KIs:

```
-----
```

Beobachtete Prävalenz	0,05 (0,02; 0,12)
Wahre Prävalenz	0,10 (0,05; 0,18)
Sensitivität	0,44 (0,14; 0,79)
Spezifität	0,99 (0,93; 1,00)
Positiver prädiktiver Wert	0,80 (0,28; 0,99)
Negativer prädiktiver Wert	0,94 (0,87; 0,98)
Positiver Wahrscheinlichkeitsquotient	36,44 (4,55; 291,84)
Negativer Wahrscheinlichkeitsquotient	0,56 (0,31; 1,01)

```
-----
```

KvarQ Pyrazinamid Vorhersage

```
> dat <- as.table(matrix(c(2,3,7,79), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
KvarQ +	2	3	5
KvarQ -	7	79	86
Total	9	82	91

E Berechnung von Sensitivität und Spezifität

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,05 (0,02; 0,12)
Wahre Prävalenz	0,10 (0,05; 0,18)
Sensitivität	0,22 (0,03; 0,60)
Spezifität	0,96 (0,90; 0,99)
Positiver prädiktiver Wert	0,40 (0,05; 0,85)
Negativer prädiktiver Wert	0,92 (0,84; 0,97)
Positiver Wahrscheinlichkeitsquotient	6,07 (1,16; 31,68)
Negativer Wahrscheinlichkeitsquotient	0,81 (0,57; 1,15)

Mykrobe Predictor TB Pyrazinamid Vorhersage

Nicht verfügbar

PhyResSE Pyrazinamid Vorhersage

```
> dat <- as.table(matrix(c(6,3,3,79), nrow = 2, byrow = TRUE))
> rval <- epi.tests(dat, conf.level = 0.95)
> print(rval); summary(rval)
```

	DST +	DST -	Total
PhyResSE +	6	3	9
PhyResSE -	3	79	82
Total	9	82	91

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,10 (0,05; 0,18)
Wahre Prävalenz	0,10 (0,05; 0,18)
Sensitivität	0,67 (0,30; 0,93)
Spezifität	0,96 (0,90; 0,99)
Positiver prädiktiver Wert	0,67 (0,30; 0,93)
Negativer prädiktiver Wert	0,96 (0,90; 0,99)

E.3 Sensitivität und Spezifität der einzelnen Prgogramme im Vergleich zur phänotypischen DST

Positiver Wahrscheinlichkeitsquotient 18,22 (5,47; 60,68)
Negativer Wahrscheinlichkeitsquotient 0,35 (0,14; 0,87)

TBProfiler Pyrazinamid Vorhersage

```
> dat <- as.table(matrix(c(4,3,5,79), nrow = 2, byrow = TRUE))  
> rval <- epi.tests(dat, conf.level = 0.95)  
> print(rval); summary(rval)
```

	DST +	DST -	Total
TBProfiler +	4	3	7
TBProfiler -	5	79	84
Total	9	82	91

Punktschätzer und 95% KIs:

Beobachtete Prävalenz	0,08 (0,03; 0,15)
Wahre Prävalenz	0,10 (0,05; 0,18)
Sensitivität	0,44 (0,14; 0,79)
Spezifität	0,96 (0,90; 0,99)
Positiver prädiktiver Wert	0,57 (0,18; 0,90)
Negativer prädiktiver Wert	0,94 (0,87; 0,98)
Positiver Wahrscheinlichkeitsquotient	12,15 (3,21; 45,90)
Negativer Wahrscheinlichkeitsquotient	0,58 (0,32; 1,04)

E.3 Sensitivität und Spezifität der einzelnen Prgogramme im Vergleich zur phänotypischen DST

Die Hauptaussagen aus Kapitel 4.3 bleiben bestehen. Abweichungen von über 10 % zur Tabelle 4.5 des Haupttextes sind in Rot (Wert verschlechtert) markiert.

Die Sensitivität für die INH-Resistenzvorhersage bezüglich der phänotypischen DST-Ergebnisse war für alle Programme geringer im Vergleich zur Sensitivität basierend auf den Sanger-Ergebnissen. Für die EMB-Resistenzvorhersage haben sich die Sensitivitäten bezüglich der phänotypischen DST-Ergebnisse von PhyResSE und TBProfiler verschlechtert, während für CASTB und Mykrobe Predictor TB eine Verbesserung der Sensitivität beobachtet werden konnte. Es traten ebenfalls größere Abweichungen bei der PZA-Resistenzvorhersage auf. Es konnte eine prozentuale Verschlechterung der Sensitivität für die PZA-Resistenzvorhersage von KvarQ und TBProfiler von 50 % beziehungsweise 20 % beobachtet werden.

E.4 Vorab veröffentlichte Sensitivitäten und Spezifitäten der automatisierten Programme

Tabelle E.1: Sensitivität und Spezifität der Resistenzvorhersage verglichen mit den phänotypischen DST-Ergebnissen für die einzelnen Antibiotika im Test. Sensitivität und Spezifität als Schätzer mit zugehörigem 95 % Konfidenzintervall. Berechnet mit R Version 3.1.2. Prozentuale Abweichung im Vergleich zur Tabelle 4.5 über 10 % in rot markiert.

Antibiotikum	Sanger		CASTB		KvarQ		Mykrobe Predictor TB		PhyResSE		TBProfiler	
	#R	#S	Sens	Spez	Sens	Spez	Sens	Spez	Sens	Spez	Sens	Spez
INH	29	62	0,83 (0,64;0,94)	0,98 (0,91; 1,00)	0,79 (0,60; 0,92)	0,98 (0,91; 1,00)	0,79 (0,60; 0,92)	0,95 (0,87; 0,99)	0,93 (0,77; 0,99)	0,97 (0,89; 1,00)	0,90 (0,73; 0,98)	0,84 (0,72; 0,92)
RMP	14	77	1,00 (0,68; 1,00)	0,96 (0,89; 0,99)	0,93 (0,66; 1,00)	0,95 (0,87; 0,99)	1,00 (0,68; 1,00)	0,94 (0,85; 0,98)	1,00 (0,68; 1,00)	0,94 (0,85; 0,98)	1,00 (0,68; 1,00)	0,94 (0,85; 0,98)
SM	37	54	0,30 (0,16; 0,47)	1,00 (0,90; 1,00)	0,57 (0,39; 0,73)	1,00 (0,90; 1,00)	0,57 (0,39; 0,73)	1,00 (0,90; 1,00)	0,84 (0,68; 0,94)	0,98 (0,90; 1,00)	0,57 (0,39; 0,73)	1,00 (0,90; 1,00)
EMB	14	77	0,57 (0,29; 0,82)	1,00 (0,93; 1,00)	0,50 (0,23; 0,77)	0,99 (0,93; 1,00)	0,50 (0,23; 0,77)	0,99 (0,93; 1,00)	0,86 (0,57; 0,98)	0,97 (0,91; 1,00)	0,86 (0,57; 0,98)	0,96 (0,89; 0,99)
PZA	9	82	0,44 (0,14; 0,79)	0,99 (0,93; 1,00)	0,22 (0,03; 0,60)	0,96 (0,90; 0,99)	0,22 (0,03; 0,60)	n.v.	0,67 (0,30; 0,93)	0,96 (0,90; 0,99)	0,44 (0,14; 0,79)	0,96 (0,90; 0,99)

Abkürzungen: INH-Isoniazid, RMP-Rifampicin, SM-Streptomycin, EMB-Ethambutol, PZA-Pyrazinamid, Sens-Sensitivität (richtig positive Rate); Spez-Spezifität (richtig negativ Rate); n.v.-nicht verfügbar

Tabelle E.2: Vorab publizierte Sensitivitäten und Spezifitäten der automatisierten Programme. Basierend auf dem Vergleich mit der phänotypischen Resistenztestung

Antibiotic	KvarQ		Mykrobe Predictor TB		TBProfiler	
	Sens	Spez	Sens	Spez	Sens	Spez
INH	84,50 ↓	89,90 ↓	92,00 ↑	98,8 ↓	92,80 ←	100,0 ↑
RMP	90,45 ↓	99,15 ↓	91,45 ↓	99,00 ←	96,20 ↓	98,10 ↓
SM	68,55 ↑	94,85 ↓	69,15 ↑	99,65 ←	88,70 ↓	81,70 ↓
EMB	76,50 ↑	96,05 ↓	67,40 ↑	96,90 ↓	87,1 ↑	89,70 ↓
PZA	n.v.	n.v.	n.v.	n.v.	70,9 ↑	93,90 ↓

F Untersuchung der seriellen NGS-Daten aus Usbekistan

F.1 Patient A

Patient A wurde initial gegen INH, RMP, EMB, SM und PZA resistent getestet. Die zweite Probe wurde mittels phänotypischer DST als nicht mehr EMB-resistent mit zusätzlicher OFX-Resistenz klassifiziert. Das phänotypische Resistenzprofil der dritten Probe war wieder identisch zur ersten Probe (Tabelle F.1).

Tabelle F.1: Entwicklung des phänotypischen Resistenzprofils der Probe von Patient A

Probe	INH	EMB	RMP	SM	PZA	OFX	CPM	AMK
1	R	R	R	R	R	S	S	S
2	R	S	R	R	R	R	S	S
3	R	R	R	R	R	S	S	S

Änderungen des Resistenzstatus in Grau. Abkürzungen: INH-Isoniazid, RMP-Rifampicin, EMB-Ethambutol, PZA-Pyrazinamid, SM-Streptomycin, OFX-Ofloxacin, CPM-Capreomycin, AMK-Amikacin, R-resistent, S-sensibel

Die Resistenzentwicklung im Laufe der Behandlung anhand der NGS-Daten ist in Abbildung F.1 dargestellt. Die Untersuchung der NGS-Daten ergab, dass alle drei NGS-Proben eine Mutation für EMB-Resistenz trugen. Bei den NGS-Proben 1 und 3 handelte es sich um die Mutation *embB* M406I (Position 4247431 G → A), die für die EMB-Resistenz verantwortlich war. Bei der Probe 2 wurde die Mutation *embB* M306V (Position 42474429 A → G) detektiert. Es konnten demnach unterschiedliche Mutationen für dasselbe Antibiotikum zu unterschiedlichen Zeitpunkten detektiert werden. Ebenso unterschieden sich die resistenzvermittelnden Mutationen für die PZA-Resistenz zwischen den NGS-Proben. NGS-Probe 1 trug

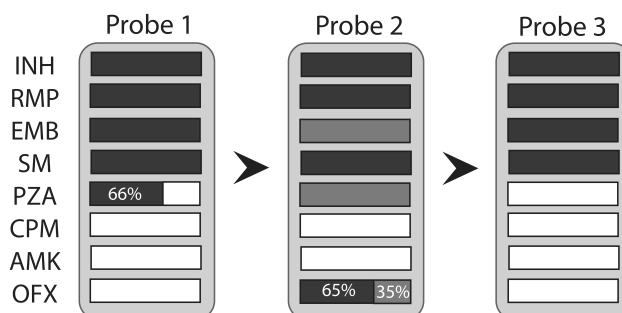


Abbildung F.1: Entwicklung des Resistenzprofils von Patient A anhand von NGS-Daten. Dunkelgrau = Resistenz-SNP 1, Grau = Resistenz-SNP 2, Weiß = kein SNP

die Mutation *pncA* C138W (Position 2288830 T → C), während Probe 2 die Mutation *pncA* Q10R (Position 2289213 T → G) aufwies. In der dritten Probe konnte an den bekannten PZA-resistenzvermittelnden Positionen (Anhang D) keine Mutation festgestellt werden. Insbesondere konnte in Probe 3 keine der zuvor detektierten Mutationen der anderen NGS-Proben detektiert werden. Die NGS-Probe 2 war die einzige, die OFX-resistent getestet wurde. Dies konnte mittels NGS-Analyse bestätigt werden. Die Probe 2 trug als einzige Probe eine resistenzvermittelnde Mutation gegen OFX. Es konnten sogar zwei unterschiedliche SNPs detektiert werden. An der Position 7570 zeigten 36 % der ausgerichteten NGS-Sequenzfragmente das mutierte T (*gyrA* A90V) und an Position 7582 zeigten ungefähr 65 % der ausgerichteten NGS-Sequenzfragmente die mutierte Base G (*gyrA* D94G). Eine Überprüfung der Genomähnlichkeiten aller drei NGS-Proben zeigte, dass sich NGS-Probe 2 von den anderen beiden NGS-Proben in 18 individuellen SNPs unterschied (Abbildung F.2). Da es sich bei allen drei NGS-Proben um Stämme derselben Linie (Beijing/2.2.1) handelte, stimmte der Großteil der SNPs (n = 815) überein und der Mischungsdetektor hat keine Mischung detektiert.

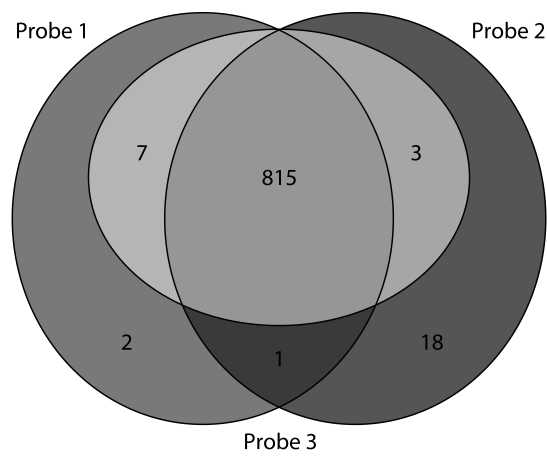


Abbildung F.2: Venn-Diagramm der SNP-Positionen der VCF-Dateien der drei NGS-Proben für Patient A. Probe 2 zeigte 18 individuelle SNPs, die in keiner der anderen NGS-Proben detektiert wurden.

F.2 Patient B

Tabelle F.2: Entwicklung des phänotypischen Resistenzprofils der Probe von Patient B

Probe	INH	EMB	RMP	SM	PZA	OFX	CPM	AMK
1	R	R	R	R	R	S	R	R
2	R	R	R	R	R	S	R	R
3	R	R	R	R	R	S	R	R
4	R	R	R	R	R	R	R	R
5	R	R	R	R	R	R	R	R

Änderungen des Resistenzstatus in Grau. Abkürzungen: INH-Isoniazid, RMP-Rifampicin, EMB-Ethambutol, PZA-Pyrazinamid, SM-Streptomycin, OFX-Ofloxacin, CPM-Capreomycin, AMK-Amikacin, R-resistent, S-sensibel

Bei Patient B war das Resistenzprofil einheitlich. Ab der vierten Probe kam eine OFX-Resistenz dazu. Anhand der NGS-Daten konnte allerdings eine komplexere Resistenzentwicklung beobachtet werden. Die NGS-Proben 1 bis 3 zeigten ein identisches Resistenzprofil und die Genome der drei NGS-Proben waren größtenteils identisch (Abbildung F.3). In NGS-Probe 1 konnten zwei unterschiedliche PZA resistenzvermittelnde SNPs beobachtet werden, bei denen jeweils 30 % der ausgerichteten NGS-Sequenzfragmente die mutierter Base zeigten. Es handelte sich um die Mutationen *pncA* F13S und den intergenischen Austausch an Position 2289252 von T → C. In Probe 2 und 3 wurde nur die *pncA* F13S Mutation mit 100 % alternativem Allel detektiert. Die Probe 4 wurde von dem Mischungsdetektor als Mischung identifiziert. Die Häufungsschwerpunkte lagen bei ungefähr 20 % und 80 % alternativem Allel. Eine Überprüfung der linienspezifischen SNPs ergab, dass Probe 1, 2, 3 und 5 der Linie 4.8 zugeordnet werden konnten. Die Probe 4 zeigte bei der Untersuchung der linienspezifischen Positionen heterogene Positionen; beijingspezifische SNPs (Linie 2.2.1) wurden mit etwa 80 % alternativem Allel detektiert und 4.8-spezifische SNPs wurden gleichzeitig mit ungefähr 20 % alternativem Allel detektiert. Auch die Frequenzen der alternativen Allele der Resistenz-SNPs untermauerten diese Hypothese. In Abbildung 5.11 ist die SNP-Verteilung dargestellt. Es konnten zwei verschiedene INH resistenzvermittelnde SNPs beobachtet werden; einen (*fabG1* -15 C → T; Position 1673425) mit ungefähr 15 % mutierter Base und einen (*katG* S315T) mit ungefähr 75 % mutierter Base. Zusätzlich gab es zwei PZA resistenzvermittelnde SNPs, die eben-

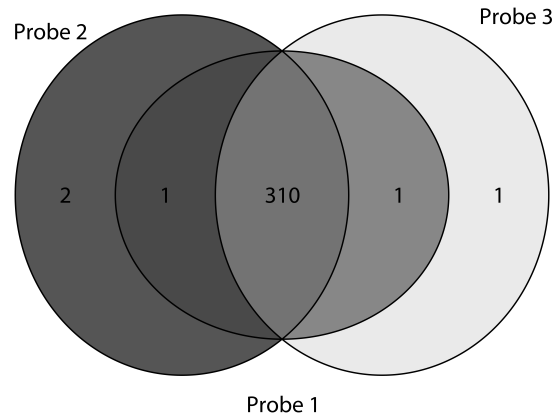


Abbildung F.3: Venn-Diagramm der SNP-Positionen der VCF-Dateien der ersten drei NGS-Proben von Patient B. Die NGS-Proben unterschieden sich in weniger als 2 SNPs voneinander.

falls mit 15 % (*pncA* F13S) beziehungsweise 75 % mutierten Basen (*pncA* I133T) detektiert wurden. Die Mutation *rrs* 1401 A → G, die für die CPM- und AMK-Resistenz der NGS-Proben 1 bis 3 verantwortlich war, wurde nur noch von etwa 15 % der NGS-Sequenzfragmente vorhergesagt. Außerdem wurde die Mutation *gyrA* D94H, die für die OFX-Resistenz verantwortlich gemacht wird, mit einer Frequenz von 15 % alternativem Allel detektiert. Mittels NGS-Analyse konnte lediglich die Probe 4 als SM-resistent bestimmt werden. Die korrespondierende Mutation *rpsL* K43R wurde mit etwa 75 % alternativem Allel detektiert. Die Mutationen für RMP- (*rpoB* S450L) und EMB-Resistenz (*embB* M306V) wurden von 100 % der ausgerichteten NGS-Sequenzfragmente vorhergesagt. In Probe 5 glich das NGS-Resistenzprofil wieder dem von Probe 1 bis 3, mit einer zusätzlichen OFX-Resistenz.

F.3 Patient C

Patient C wurden vier Proben in einem Zeitraum von acht Monaten entnommen. Die phänotypischen DST-Ergebnisse sind in Tabelle F.3 aufgetragen. Der Patient war in der initialen Probe resistent gegen alle getesteten Antibiotika bis auf OFX. Die Probe 2 war laut phänotypischer DST nicht mehr EMB resistent, Probe 3 hatte das identische Resistenzprofil wie Probe 1 und die letzte Probe war abermals nicht mehr EMB-resistent und nicht mehr RMP-resistent. Das heißt, der Patient war laut Definition ab Probe 4 kein MDR-TB Patient mehr. Zusätzlich kam in Probe 4 eine OFX-Resistenz hinzu.

Die NGS-Analyse der Sequenzdaten der vier Proben hat ergeben, dass alle NGS-

Tabelle F.3: Entwicklung des phänotypischen Resistenzprofils der Probe von Patient C

Probe	INH	EMB	RMP	SM	PZA	OFX	CPM	AMK
1	R	R	R	R	R	S	R	R
2	R	S	R	R	R	S	R	R
3	R	R	R	R	R	S	R	R
4	R	S	S	R	S	R	R	R

Änderungen des Resistenzstatus in Grau. Abkürzungen: INH-Isoniazid, RMP-Rifampicin, EMB-Ethambutol, PZA-Pyrazinamid, SM-Streptomycin, OFX-Ofloxacin, CPM-Capreomycin, AMK-Amikacin, R-resistent, S-sensibel

Proben dieselbe *embB* M306V Mutation trugen und somit gegen EMB resistent waren. Eine OFX-Resistenz konnte mittels NGS-Analyse für keine der NGS-Proben festgestellt werden. Über die Dauer der Behandlung haben sich die Mutationen *rrs* 514 A → C (SM), *pncA* Q141P (PZA) und *rrs* 1401 A → G (AMK/CPM) in der Population durchgesetzt (Abbildung 5.16). Ein gegenteiliges Ereignis zeigte die RMP resistenzvermittelnde Mutation *rpoB* H445R. In den NGS-Daten der Probe 1 war diese Mutation bei 16 % der ausgerichteten NGS-Sequenzfragmente zu beobachten, in den NGS-Daten der Probe 2 bei 10 %, in den NGS-Daten der Probe 3 bei 30 % und in den NGS-Daten der Probe 4 konnte diese Variante nicht mehr detektiert werden. Die Abbildung 5.17 zeigt, dass anhand der Genomdaten zwischen allen NGS-Proben nur wenige SNPs Unterschied beobachtet werden konnten.

F.4 Patient D

Tabelle F.4: Entwicklung des phänotypischen Resistenzprofils der Probe von Patient D

Probe	INH	EMB	RMP	SM	PZA	OFX	CPM	AMK
1	R	S	R	R	S	S	S	S
2	R	R	R	R	R	S	S	S
3	R	R	R	R	R	R	S	S

Änderungen des Resistenzstatus in Grau. Abkürzungen: INH-Isoniazid, RMP-Rifampicin, EMB-Ethambutol, PZA-Pyrazinamid, SM-Streptomycin, OFX-Ofloxacin, CPM-Capreomycin, AMK-Amikacin, R-resistent, S-sensibel

Patient D wurde initial mittels phänotypischer DST gegen INH, RMP und SM resistent getestet (DST; Tabelle F.4). Probe 2 zeigte bei der phänotypischen DST weitere Resistenzen gegen EMB und PZA (Tabelle F.4). Die letzte Probe war zusätzlich gegen OFX resistent.

Die NGS-Analyse hat ergeben, dass die erste Probe anhand der NGS-Daten einem Stamm der Linie 4.8 zugewiesen werden konnte, während die NGS-Proben 2 und 3 einem Stamm der Linie Beijing angehörten (Linie 2.2.1). Es handelte sich also um unterschiedliche Stämme. Abbildung 5.13 zeigt die Genomähnlichkeiten der drei NGS-Proben anhand eines Venn-Diagramms. In Probe 1 konnten 801 individuelle SNPs detektiert werden, die diese NGS-Probe von den anderen NGS-Proben abgrenzte. Dieser Unterschied zeigte sich ebenfalls bei den Resistenz-SNPs. Mit Hilfe der NGS-Analyse konnte gezeigt werden, dass alle NGS-Proben INH-, EMB-, RMP- und SM-resistent waren. Die Resistenzentwicklung über die Zeit anhand der NGS-Daten ist in Abbildung 5.12 dargestellt. Die Mutationen in den NGS-Daten, die für INH-, EMB- und RMP-Resistenz verantwortlich waren, waren für alle drei NGS-Proben die gleichen (*katG* S315T, *rpoB* S450L und *embB* M306V). Bei diesen Mutationen handelt es sich allerdings um sehr häufig vorkommende Mutationen. Bei der SM-Resistenz trat ein Unterschied auf. In den NGS-Daten der Probe 1 (Linie 4.8) konnte die Mutation *rrs* 514 A → C beobachtet werden, während in den NGS-Daten der Proben 2 und 3 (Beijing/Linie 2.2.1) die Mutation *rpsL* L43R als verantwortliche Mutation identifiziert werden konnte. In den NGS-Daten der Probe 3 konnte zusätzlich die Mutation *gyrA* A90V, die für die OFX-Resistenz verantwortlich gemacht wird, detektiert werden.

F.5 Patient E

Der Patient E war laut phänotypischer DST initial gegen INH, RMP, SM und PZA resistent (Tabelle F.5). Das phänotypische DST-Ergebnis war für Probe 2 zusätzlich EMB resistent und PZA sensibel. Probe 3 wurde mittels phänotypischer DST gegen INH, RMP, EMB, SM, PZA und zusätzlich OFX resistent getestet.

Die NGS-Analyse zeigte, dass alle drei NGS-Proben dieselben Mutationen für

Tabelle F.5: Entwicklung des phänotypischen Resistenzprofils der Proben von Patient E

Probe	INH	EMB	RMP	SM	PZA	OFX	CPM	AMK
1	R	S	R	R	R	S	S	S
2	R	R	R	R	S	S	S	S
3	R	R	R	R	R	R	S	S

Änderungen des Resistenzstatus in Grau. Abkürzungen: INH-Isoniazid, RMP-Rifampicin, EMB-Ethambutol, PZA-Pyrazinamid, SM-Streptomycin, OFX-Ofloxacin, CPM-Capreomycin, AMK-Amikacin, R-resistent, S-sensibel

INH-, EMB-, RMP-, SM- und PZA-Resistenz trugen und folglich resistent waren. Bei den Mutationen handelte es sich um *katG* S315T (INH), *embB* M306V (EMB), *rpoB* S450L (RMP), *rpsL* L43R (SM) und *pncA* Q10P (PZA). Alle Mutationen wurden von 100 % der ausgerichteten NGS-Sequenzfragmente an den jeweiligen Positionen vorhergesagt. NGS-Probe 3 trug die zusätzliche Mutation *gyrA* A90V, die für die OFX-Resistenz in NGS-Probe 3 verantwortlich war. Die mutierte Base (Position 7570 C → T) wurde von 56 % der ausgerichteten NGS-Sequenzfragmente vorhergesagt. Die Resistenzentwicklung über die Zeit ist in Abbildung F.5 dargestellt. Die Genome der drei NGS-Proben waren sehr ähnlich (Abbildung F.4). Die NGS-Daten der Proben 1 und 3 teilten zehn SNPs, die in den NGS-Daten der Probe 2 nicht detektiert werden konnten.

F.6 Patient F

Patient F war laut phänotypischer DST initial gegen INH, EMB, RMP und SM resistent. Die Entwicklung des Resistenzprofils mittels phänotypischer DST ist in

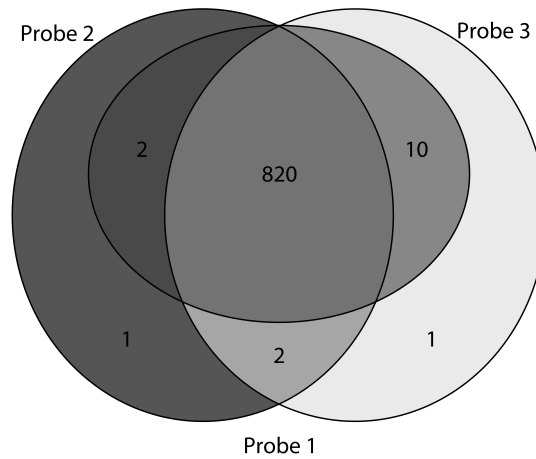


Abbildung F.4: Venn-Diagramm der SNP-Positionen der VCF-Dateien von Patient E. Es konnten nur wenige SNPs Unterschied beobachtet werden.

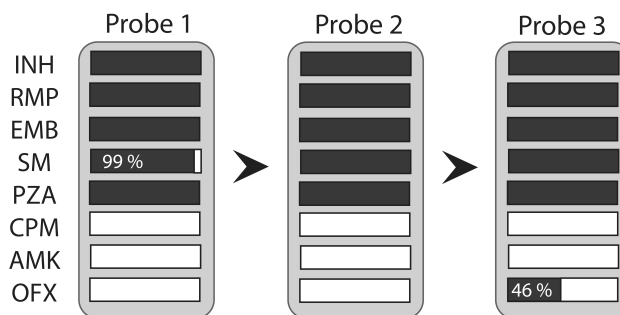


Abbildung F.5: Entwicklung des Resistenzprofils von Patient E anhand von NGS-Daten. Grau = Resistenz-SNP, Weiß = kein SNP

Tabelle F.6 dargestellt. Im Vergleich zur Probe 1 konnten für Probe 2 zusätzliche Resistenzen gegen PZA und OFX beobachtet werden. Die Probe 4 wurde mittels phänotypischer DST zusätzlich CPM resistent getestet. Das phänotypische Resistenzprofil ließ auf eine Infektion mit einem Stamm der Linie Beijing schließen, der im Laufe der Behandlung zusätzliche Resistenzen gegen PZA, OFX und CPM entwickelt hat.

Mit Hilfe der NGS-Analyse wurden für jede der vier NGS-Proben dieselben Re-

Tabelle F.6: Entwicklung des phänotypischen Resistenzprofils der Proben von Patient F

Probe	INH	EMB	RMP	SM	PZA	OFX	CPM	AMK
1	R	R	R	R	S	S	S	S
2	R	R	R	R	R	R	S	S
3	R	R	R	R	R	R	S	S
4	R	R	R	R	R	R	R	S

Änderungen des Resistenzstatus in Grau. Abkürzungen: INH-Isoniazid, RMP-Rifampicin, EMB-Ethambutol, PZA-Pyrazinamid, SM-Streptomycin, OFX-Oloxacin, CPM-Capreomycin, AMK-Amikacin, R-resistent, S-sensibel

sistenzmutationen für INH, EMB, RMP, SM und PZA detektiert (Abbildung F.6).

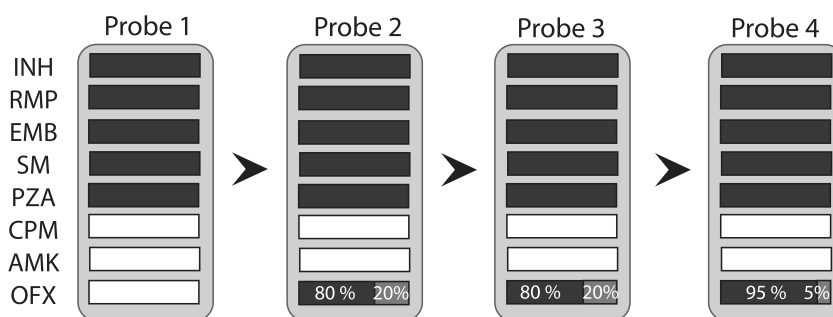


Abbildung F.6: Entwicklung des Resistenzprofils von Patient F anhand von NGS-Daten. Dunkelgrau = Resistenz-SNP 1, Grau = Resistenz-SNP 2, Weiß = kein SNP

Die Mutationen *katG* S315T (INH), *embB* M306V (EMB), *rpoB* S450L (RMP), *rpsL* L43Z (SM) und *pncA* Q10P (PZA) konnten jeweils bei 100% der ausgegerichteten NGS-Sequenzfragmente detektiert werden. Ab NGS-Probe 2 konnte

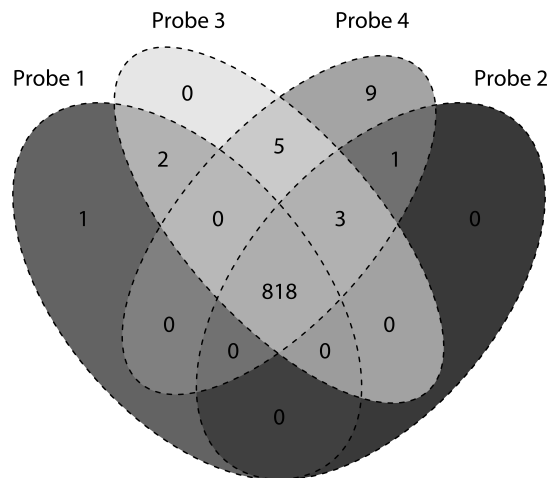


Abbildung F.7: Venn-Diagramm der SNP-Positionen der VCF-Dateien der NGS-Proben von Patient F. Die NGS-Probe 4 war neun SNPs von den anderen Proben entfernt, die nur geringe Unterschiede aufwiesen.

eine OFX-Resistenz festgestellt werden. In den NGS-Daten der Proben 2 und 3 konnten jeweils zwei mögliche Kandidatenmutationen im Verhältnis 8:2 detektiert werden. Dabei wurde die Mutation *gyrA* A90V von ungefähr 20% der ausgerichteten NGS-Sequenzfragmente und die Mutation *gyrA* D94G von jeweils ungefähr 80% der ausgerichteten NGS-Sequenzfragmente vorhergesagt. In den NGS-Daten der Probe 3 konnten ebenfalls beide Mutationen beobachtet werden, allerdings wurde die Mutation *gyrA* A90V nur noch bei 5% der ausgerichteten NGS-Sequenzfragmente detektiert.

Der Mischungsdetektor hat keine Mischinfektion detektiert, die zwei Mutation für das gleiche Antibiotikum mit den beobachteten Frequenzen der alternativen Allele erklären würde. Vielmehr ähnelten sich die Genome der vier NGS-Proben derart, dass von demselben Stamm ausgegangen werden konnte (Abbildung F.7).

F.7 Patient G

Patient G wurde mittels phänotypischer DST initial für INH, RMP, EMB, SM und PZA resistent getestet. In der zweiten Probe hatte sich das phänotypische Resistenzprofil nicht verändert. Ab Probe 3 konnte eine zusätzliche Resistenz gegen OFX beobachtet werden.

Tabelle F.7: Entwicklung des phänotypischen Resistenzprofils der Proben von Patient G

Probe	INH	EMB	RMP	SM	PZA	OFX	CPM	AMK
1	R	R	R	R	R	S	S	S
2	R	R	R	R	R	S	S	S
3	R	R	R	R	R	R	S	S

Änderungen des Resistenzstatus in Grau. Abkürzungen: INH-Isoniazid, RMP-Rifampicin, EMB-Ethambutol, PZA-Pyrazinamid, SM-Streptomycin, OFX-Ofloxacin, CPM-Capreomycin, AMK-Amikacin, R-resistent, S-sensibel

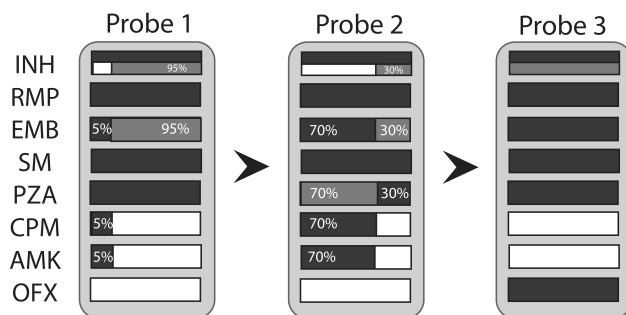


Abbildung F.8: Entwicklung des Resistenzprofils von Patient G anhand von NGS-Daten. Dunkelgrau = Resistenz-SNP 1, Grau = Resistenz-SNP 2, Weiß = kein SNP

Mit Hilfe der NGS-Daten von Patient G konnte eine sehr viel komplexere SNP-Verteilung dargestellt werden, als das phänotypische Resistenzprofil vermuten ließ. In Abbildung F.8 ist dargestellt, wie sich die Mutationen im Laufe der Behandlung veränderten. In den NGS-Daten der Probe 1 wurden zwei INH-resistenzvermittelnde Mutationen detektiert. Die Mutation *katG* S315T wurde bei 100 % der ausgerichteten NGS-Sequenzfragmente detektiert. Die Mutation *fabG1*

-8 T → C wurde bei ungefähr 95 % der ausgerichteten NGS-Sequenzfragmente beobachtet. Für EMB-Resistenz konnten ebenfalls zwei Mutationen in den NGS-Daten der Probe 1 detektiert werden. Die Mutation *embB* M306I wurde von 5 % der ausgerichteten NGS-Sequenzfragmente vorhergesagt, während die Mehrheit der ausgerichteten NGS-Sequenzfragmente (95 %) die Mutation *embB* M306V vorhersagten. Außerdem konnte eine Resistenzmutation (*rrs* 1401 A → G) für AMK/CPM detektiert werden. Diese war in den NGS-Daten der Probe 1 bei 5 % der ausgerichteten NGS-Sequenzfragmente vorhanden. Die Mutationen *rpoB* S450L (RMP), *rpsL* K43R (SM) und *pncA* Q10P (PZA) wurden von 100 % der ausgerichteten NGS-Sequenzfragmente vorhergesagt.

In den NGS-Daten der Probe 2 konnten die gleichen Mutationen mit einem veränderten Verhältnis beobachtet werden (Abbildung F.8). Die zweite INH-resistenzvermittelnde Mutation *fabG1* -8 T → C war in 30 % der ausgerichteten NGS-Sequenzfragmente vorhanden. Ebenso war die EMB-resistenzvermittelnde Mutation, die in der NGS-Probe 1 zur Hauptpopulation zählte, in etwa 30 % der ausgerichteten NGS-Sequenzfragmente vorhanden und die zweite Mutation (*embB* M306I), die in den NGS-Daten der Probe 1 bei 5 % der ausgerichteten NGS-Sequenzfragmente detektiert wurde, war bei 70 % der ausgerichteten NGS-Sequenzfragmente zu beobachten. Für PZA-Resistenz gab es im selben Verhältnis (7:3) zwei unterschiedliche Mutationen an der Genomposition 2155168 (*pncA* Q10P ≐ 30 % und *pncA* Q10R ≐ 70%). Die AMK/CPM-resistenzvermittelnde Mutation wurde bei 70 % der ausgerichteten NGS-Sequenzfragmente detektiert. Das Venn-Diagramm der SNPs aus den NGS-Daten der vier Proben von Patient G zeigte nur geringe SNP-Unterschiede zwischen den NGS-Proben (Abbildung F.9).

F.8 Patient H

Patient H wurde mittels phänotypischer DST initial gegen alle aufgeführten Antibiotika außer EMB und OFX resistent getestet (Tabelle F.8). Die Probe 2 von Patient H war laut phänotypischer DST nicht mehr AMK/CPM resistent. Probe 3 war hingegen zusätzlich gegen OFX und CPM resistent, allerdings fehlte die AMK-Resistenz der initialen Probe. Probe 4 war laut phänotypischer DST gegen alle aufgeführten Antibiotika bis auf AMK/CPM resistent.

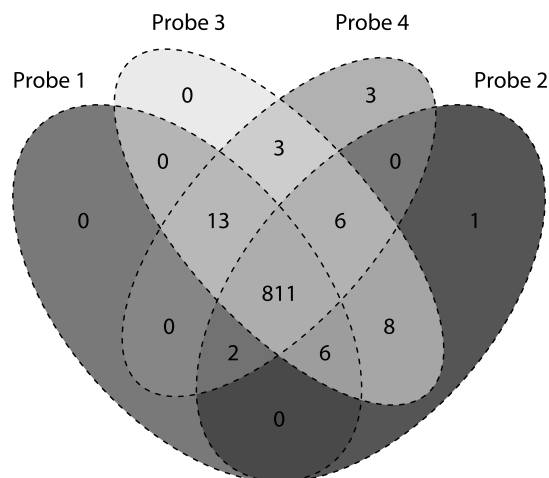


Abbildung F.9: Venn-Diagramm der SNP-Positionen der VCF-Dateien von Patient G. Die NGS-Probe 4 unterschied sich in 9 individuellen SNPs von den anderen Proben, die nur geringe Unterschiede aufwiesen.

Der Fall von Patient H wurde bereits als Reinfektion mit einem zweiten Stamm in der Studie von Merker *et al.* im Jahr 2013 identifiziert [139]. Für die Studie von Merker *et al.* wurden ebenfalls NGS-Daten verwendet. Die Analyseprogramme, sowie die Abfolge unterschieden sich allerdings von der für diese Arbeit verwendeten Analyseabfolge (Abbildung 2.3). Dennoch bestätigen die Analysemethoden, die in dieser Arbeit verwendet und etabliert wurden, die Ergebnissen von Merker *et al.*.

Tabelle F.8: Entwicklung des phänotypischen Resistenzprofils der Proben von Patient H

Probe	INH	EMB	RMP	SM	PZA	OFX	CPM	AMK
1	R	S	R	R	R	S	R	R
2	R	S	R	R	R	S	S	S
3	R	S	R	R	R	R	R	S
4	R	R	R	R	R	R	S	S

Änderungen des Resistenzstatus in Grau. Abkürzungen: INH-Isoniazid, RMP-Rifampicin, EMB-Ethambutol, PZA-Pyrazinamid, SM-Streptomycin, OFX-Ofloxacin, CPM-Capreomycin, AMK-Amikacin, R-resistent, S-sensibel

Die Resistenzentwicklung der Proben anhand der NGS-Daten ist in Abbildung F.10 dargestellt. Die NGS-Probe 1 war eindeutig INH-, RMP- und SM-resistent. Die verantwortlichen Mutationen *katG* S315T, *rpoB* S450L und *rpsL* K43R wurden jeweils von 100 % der ausgerichteten NGS-Sequenzfragmente vorhergesagt. Für EMB und PZA konnten jeweils zwei unterschiedliche Mutationen im Verhältnis 6:4 beobachtet werden.

Im Zusammenhang mit EMB standen die Mutationen *embA* -12 C → T (40 %) und *embB* M306V (60 %). Im Zusammenhang mit PZA konnten die Mutationen *pncA* -11 C → G (40 %) und *pncA* Q10P (60 %) beobachtet werden. Die Mutation *rrs* 1401 A → G, die für AMK/CPM-Resistenz verantwortlich ist, wurde mit einer Frequenz von 10 % alternativem Allel detektiert.

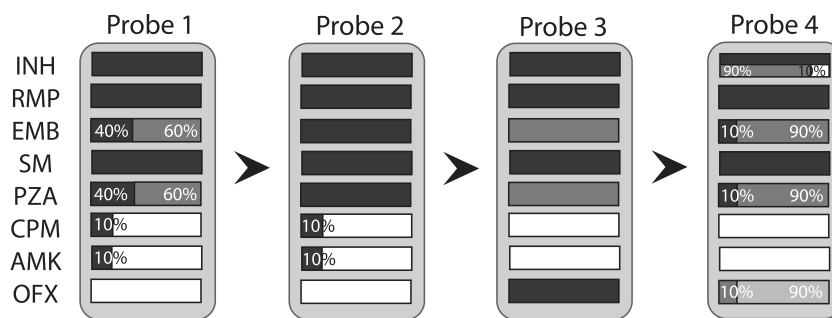


Abbildung F.10: Entwicklung des Resistenzprofils von Patient H anhand von NGS-Daten. Dunkelgrau = Resistenz-SNP 1, Grau = Resistenz-SNP 2, Hellgrau = Resistenz-SNP 3, Weiß = kein SNP

In den NGS-Daten der Probe 2 fanden sich dieselben Mutationen für INH, RMP und SM mit 100 % alternativem Allel. Zusätzlich wurden die Mutationen für EMB- und PZA-Resistenz, die in den NGS-Daten der Probe 1 bei 40 % der ausgerichteten NGS-Sequenzfragmente beobachtet werden konnten, ebenfalls von 100 % der NGS-Sequenzfragmente vorhergesagt. Die Mutation für die AMK/CPM-Resistenz blieb unverändert bei 10 % alternativem Allel. Die NGS-Probe 3 wies dieselben Mutationen für INH-, RMP- und SM-Resistenz auf. In dieser NGS-Probe fanden sich allerdings die resistenzvermittelnden Mutationen für EMB und PZA, die in den NGS-Daten der Probe 1 mit 60 % alternativem Allel vorhanden waren, mit 100 % alternativem Allel. Eine Mutation für die AMK/CPM-Resistenz konnte nicht detektiert werden. Zusätzlich befand sich eine weitere Mutation (*gyrA*

D94G) in den NGS-Daten der Probe 3, die eine OFX-Resistenz zur Folge hatte. Die Mutation wurde von 99 % der ausgerichteten NGS-Sequenzfragmente vorhergesagt.

Die NGS-Probe 4 unterschied sich anhand der Mutationen deutlich von den anderen NGS-Proben. Es gab eine zusätzliche INH-resistenzvermittelnde Mutation (*fabG1* -8 T → C), die von 90 % der ausgerichteten NGS-Sequenzfragmente vorhergesagt wurde. Es traten erneut beide resistenzvermittelnde Mutationen für EMB und PZA auf, diesmal mit jeweils 10 % beziehungsweise 90 % alternativem Allel. Die OFX-resistenzvermittelnde Mutation aus der NGS-Probe 3 konnte nicht mehr detektiert werden, dafür konnten zwei andere Mutationen beobachtet werden. Die eine Mutation *gyrA* D94H wurde von 10 % der ausgerichteten NGS-Sequenzfragmente vorhergesagt und die zweite *gyrA* A90V von 90 % der ausgerichteten NGS-Sequenzfragmente. In dem Venn-Diagramm der SNPs aller NGS-Proben von Patient H, konnten nur geringe Unterschiede beobachtet werden (Abbildung F.11). Die NGS-Daten der Proben 3 und 4 teilten 14 SNPs, die in den NGS-Daten der Proben 1 und 2 nicht beobachtet werden konnten und NGS-Probe 3 hatte neun individuelle SNPs.

F.9 Patient I

Patient I war laut phänotypischer DST initial gegen INH, RMP und SM resistent und entwickelte im Laufe der Behandlung Resistenzen gegen alle aufgeführten Antibiotika (Tabelle F.9). Das phänotypische Resistenzprofil zeigte eine schrittweise Entwicklung der Resistenzen.

Mit Hilfe der NGS-Analyse konnte gezeigt werden, dass in allen NGS-Proben dieselbe Mutation *rpoB* S450L verantwortlich für die RMP-Resistenz war. Anhand der Analyse der NGS-Daten der Proben 1, 3, 4 und 5 konnten zudem jeweils die Mutationen *katG* S315N (INH), *embB* M306V (EMB) und *rpsL* K88R (SM) detektiert werden. In den NGS-Daten der Probe 1 befanden sich keine weiteren resistenzvermittelnden Mutationen. In den NGS-Daten der Proben 3, 4 und 5 konnten zusätzlich die gemeinsamen Mutationen *pncA* L120R (PZA), *rrs* 1401 A → G (AMK/CPM) und *gyrA* D94N (OFX) detektiert werden. Die NGS-Daten der Probe 2 von Patient I unterschieden sich anhand von 5 Mutationen von den

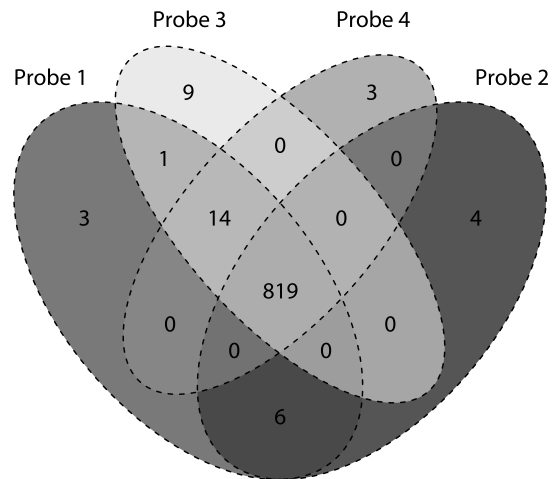


Abbildung F.11: Venn-Diagramm der SNP-Positionen der VCF-Dateien von Patient H. Es konnten nur wenige SNPs Unterschied beobachtet werden. Die NGS-Proben 3 und 4 unterschieden sich von NGS-Probe 1 und 2 mit 14 gemeinsamen SNPs.

Tabelle F.9: Entwicklung des phänotypischen Resistenzprofils der Proben von Patient I

Probe	INH	EMB	RMP	SM	PZA	OFX	CPM	AMK
1	R	S	R	R	S	S	S	S
2	R	S	R	R	S	S	R	R
3	R	S	R	R	R	R	R	R
4	R	R	R	R	R	R	R	R
5	R	R	R	R	R	R	R	R

Änderungen des Resistenzstatus in Grau. Abkürzungen: INH-Isoniazid, RMP-Rifampicin, EMB-Ethambutol, PZA-Pyrazinamid, SM-Streptomycin, OFX-Ofloxacin, CPM-Capreomycin, AMK-Amikacin, R-resistent, S-sensibel

NGS-Daten der anderen Proben (Abbildung 5.14). Die Positionen für der resistenzvermittelnden Mutationen gegen INH und EMB waren die gleichen, allerdings unterschied sich die mutierte Basen und damit der Aminosäureaustausch. In Abbildung 5.15 ist zu erkennen, dass sich die NGS-Daten der Probe 2 deutlich von den anderen Proben unterschieden. In den NGS-Daten von Probe 2 wurden 59 SNPs detektiert, die in keiner der anderen Proben auftraten. Außerdem konnten anhand der NGS-Daten der Proben 1, 3, 4 und 5 52 gemeinsame SNPs beobachtet werden, die nicht in der Probe 2 detektiert werden konnten.

In den NGS-Daten der Probe 2 konnte für die INH-Resistenz die Mutation *katG* S315T identifiziert werden. Die EMB-resistenzvermittelnde Mutation war *embB* M306L, die SM Mutation lautete *rpsL* K43R und für die PZA-Resistenz war die Mutation *pncA* D12N verantwortlich. Eine OFX-resistenzvermittelnde Mutation konnte nicht detektiert werden. Die Mutation *rrs* 1401 A → G (AMK/CPM) der NGS-Proben 3 bis 5 konnte in den NGS-Daten der Probe 2 ebenfalls detektiert werden.

Alle Proben konnten anhand der NGS-Analyse, im Gegensatz zu den phänotypischen DST-Ergebnissen, als EMB-resistent identifiziert werden.

F.10 Patient J

Nach phänotypischen DST-Ergebnissen war Patient J initial resistent gegen INH, RMP, SM, PZA, CPM und AMK. Ab der Probe 3 wurden zusätzliche Resistenzen gegen EMB und OFX detektiert, sodass Patient J laut phänotypischer DST ab Probe 3 gegen alle aufgeführten Antibiotika resistent war (Tabelle F.10).

Mit Hilfe der NGS-Analyse konnte gezeigt werden, dass in allen NGS-Proben dieselbe EMB-Resistenzmutation *embB* M306V detektiert wurden und somit alle Proben EMB-resistent waren. Außerdem konnten in den NGS-Daten aller Proben dieselben Mutationen für die weiteren Resistenzen detektiert werden. Es handelte sich um die SNPs *katG* S315T (INH), *rpoB* S450T (RMP), *rpsL* K43R (SM), *pncA* W68R (PZA), *rrs* 1401 A → G (AMK/CPM). In den NGS-Daten der Probe 3 ließ sich die Mutation *gyrA* D94G, die für die OFX-Resistenz verantwortlich gemacht wird, bei 95 % der NGS-Sequenzfragmente nachweisen. In den NGS-Daten

Tabelle F.10: Entwicklung des phänotypischen Resistenzprofils der Proben von Patient J

Probe	INH	EMB	RMP	SM	PZA	OFX	CPM	AMK
1	R	S	R	R	R	S	R	R
2	R	S	R	R	R	S	R	R
3	R	R	R	R	R	R	R	R
4	R	R	R	R	R	R	R	R
5	R	R	R	R	R	R	R	R

Änderungen des Resistenzstatus in Grau. Abkürzungen: INH-Isoniazid, RMP-Rifampicin, EMB-Ethambutol, PZA-Pyrazinamid, SM-Streptomycin, OFX-Ofloxacin, CPM-Capreomycin, AMK-Amikacin, R-resistent, S-sensibel

der Probe 4 erhöhte sich dieser Anteil auf 99 % und in den NGS-Daten der Probe 5 zeigten alle ausgerichteten NGS-Sequenzfragmente die genannte Mutation (Abbildung F.12).

Ein Vergleich der Genomähnlichkeiten zeigte sehr nah verwandte Stämme in allen fünf NGS-Proben (Abbildung F.13). Eine Ausnahme bildete die Probe 2, deren NGS-Daten sich durch zehn individuelle SNPs von den anderen NGS-Proben unterschieden.

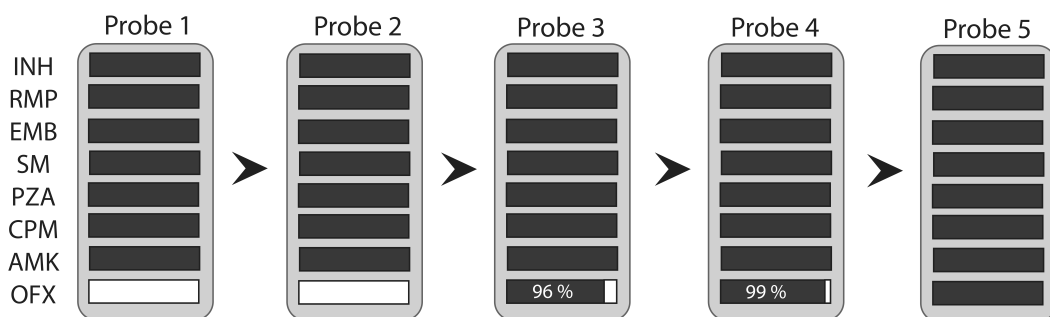


Abbildung F.12: Entwicklung des Resistenzprofils von Patient J anhand von NGS-Daten. Dunkelgrau = Resistenz-SNP, Weiß = kein SNP

F.11 Patient K

Das phänotypische Resistenzprofil von Patient K ist in Tabelle F.11 dargestellt. Initial wurden Resistenzen gegen alle aufgeführten Antibiotika mit Ausnahme

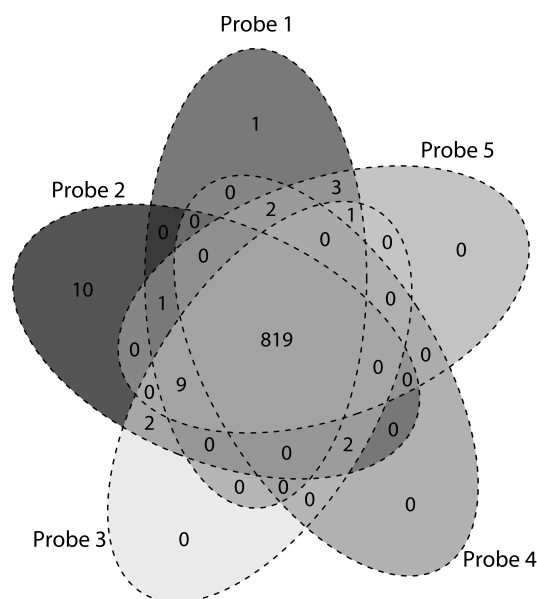


Abbildung F.13: Venn-Diagramm der SNP-Positionen der VCF-Dateien von Patient J. Alle NGS-Proben waren sehr nah verwandt. NGS-Probe 2 zeigte allerdings zehn individuelle SNPs, die in keiner anderen NGS-Proben detektiert wurden.

von PZA und OFX festgestellt. Ab Probe 2 waren die Proben zusätzlich resistent gegen PZA. Probe 3 wurde sensibel auf EMB getestet, wohingegen Probe 4 laut phänotypischer DST wieder resistent gegen EMB war. Zusätzlich wurde Probe 4 OFX-resistent getestet. Probe 5 war laut phänotypischer DST gegen alle aufgeführten Antibiotika resistent.

In Abbildung F.14 ist die Entwicklung der Resistenzmutationen der NGS-Proben von Patient K basierend auf den NGS-Analyseergebnissen dargestellt.

Bei allen NGS-Proben konnten dieselben resistenzvermittelnden Mutationen mit 100 % alternativem Allel gegen INH (*katG* S315T), RMP (*rpoB* S450L) und SM (*rpsL* K43R) detektiert werden. In den NGS-Daten der Probe 1 wurden drei verschiedenen EMB-resistenzvermittelnde Mutationen detektiert. Die Mutation *embB* L74R wurde von 70 % der ausgerichteten NGS-Sequenzfragmente vorhergesagt, während die Mutationen *embB* M306V und *embB* D328G von jeweils 50 % der ausgerichteten NGS-Sequenzfragmente vorhergesagt wurden. Eine ähnliche Verteilung konnte für die NGS-Probe 3 beobachtet werden, wobei die Mutation *embB* L74R ebenfalls von je 50 % der ausgerichteten NGS-Sequenzfragmente vorhergesagt wurde. In den NGS-Daten der Probe 2 konnte ausschließlich die

Tabelle F.11: Entwicklung des phänotypischen Resistenzprofils der Proben von Patient K

Probe	INH	EMB	RMP	SM	PZA	OFX	CPM	AMK
1	R	R	R	R	S	S	R	R
2	R	R	R	R	R	S	R	R
3	R	S	R	R	R	S	R	R
4	R	R	R	R	R	R	R	R
5	R	R	R	R	R	R	R	R

Änderungen des Resistenzstatus in Grau. Abkürzungen: INH-Isoniazid, RMP-Rifampicin, EMB-Ethambutol, PZA-Pyrazinamid, SM-Streptomycin, OFX-Ofloxacin, CPM-Capreomycin, AMK-Amikacin, R-resistent, S-sensibel

Mutation *embB* M306V mit 100 % alternativem Allel detektiert werden, während in den NGS-Daten der Proben 4 und 5 ausschließlich die anderen beiden Mutationen beobachtet werden konnten.

In den NGS-Daten der Probe 1 konnten zwei Mutationen gegen PZA detektiert werden, die jeweils mit 50 % alternativem Allel auftraten (*pncA* Q10P und *pncA* K48E). In den NGS-Daten der Probe 2 wurde ausschließlich die erste Mutation mit 100 % alternativem Allel detektiert. In den NGS-Daten der Probe 3 konnten wiederum beide Mutationen im Verhältnis 1:1 beobachtet werden und ab der NGS-Probe 4 trat nur noch die zweite Mutation auf.

Die Mutation, die in den NGS-Daten der Proben 1, 3, 4 und 5 für die AMK/CPM-Resistenz verantwortlich war, trat in den NGS-Daten der Proben 1 und 3 bei jeweils 50 % der ausgerichteten NGS-Sequenzfragmente auf. Ab NGS-Probe 4 war die Mutation in 100 % der NGS-Sequenzfragmente vorhanden.

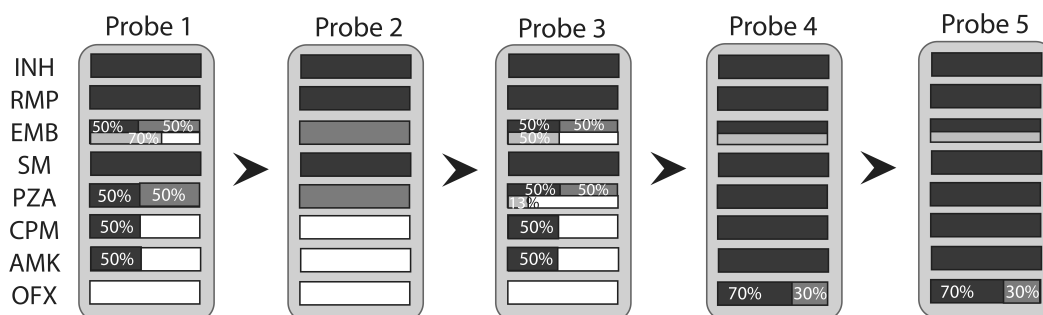


Abbildung F.14: Entwicklung des Resistenzprofils von Patient K anhand von NGS-Daten. Dunkelgrau = Resistenz-SNP 1; Grau = Resistenz-SNP 2; Hellgrau = Resistenz-SNP 3; Weiß = kein SNP

Außerdem konnten anhand der NGS-Daten der Probe 4 zwei weitere Mutationen detektiert werden, die für die OFX-Resistenz aus Tabelle F.11 verantwortlich waren. Es handelte sich um die Genomposition 7582 an der bei ungefähr 30% der ausgerichteten NGS-Sequenzfragmente die Mutation *gyrA* D94A und bei den übrigen NGS-Sequenzfragmenten die Mutation *gyrA* D94G beobachtet werden konnte.

Der Vergleich der Genomähnlichkeiten im Venn-Diagramm zeigte keine deutlichen Unterschiede zwischen den NGS-Proben (Abbildung F.15).

F.12 Patient L

Patient L wurde anhand der phänotypischen DST-Ergebnisse initial als INH-, RMP-, SM-, sowie AMK/CPM-resistent eingestuft (Tabelle F.12). In der Probe 2 konnten weitere Resistenzen gegen EMB und PZA festgestellt werden, während der Test auf CPM sensibel ausfiel. Probe 3 zeigte phänotypisch Resistenzen gegenüber allen aufgeführten Antibiotika. Probe 4 war hingegen sensibel gegenüber PZA.

Mit Hilfe der NGS-Analyse konnte gezeigt werden, dass alle Stämme die INH-Resistenzmutation *katG* S315T, die RMP-Resistenzmutation *rpoB* S450L, die SM-Resistenzmutation *rpsL* K43R, sowie die AMK/CPM-Resistenzmutation *rrs* 1401 A → G trugen (Abbildung F.16). Bei keiner der NGS-Proben konnte allerdings eine Mutation detektiert werden, die für die phänotypische PZA-Resistenz

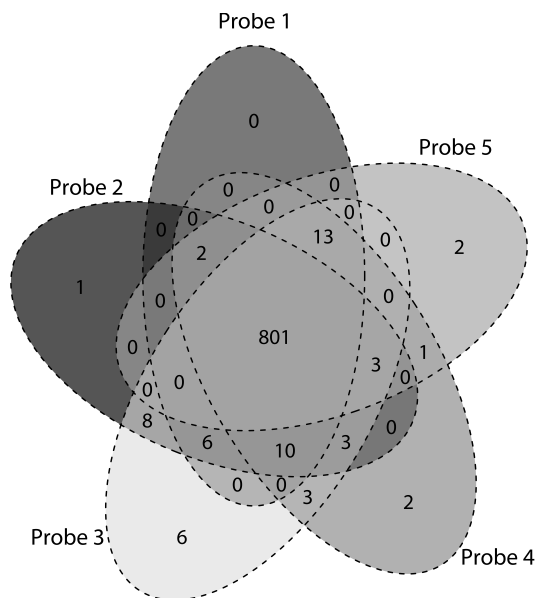


Abbildung F.15: Venn-Diagramm der SNP-Positionen der VCF-Dateien von Patient K. Alle NGS-Proben zeigen dieselben SNPs.

Tabelle F.12: Entwicklung des phänotypischen Resistenzprofils der Proben von Patient L

Probe	INH	EMB	RMP	SM	PZA	OFX	CPM	AMK
1	R	S	R	R	S	S	R	R
2	R	R	R	R	R	S	S	R
3	R	R	R	R	R	R	R	R
4	R	R	R	R	S	R	R	R

Änderungen des Resistenzstatus in Grau. Abkürzungen: INH-Isoniazid, RMP-Rifampicin, EMB-Ethambutol, PZA-Pyrazinamid, SM-Streptomycin, OFX-Ofloxacin, CPM-Capreomycin, AMK-Amikacin, R-resistent, S-sensibel

verantwortlich sein könnte. Im Zusammenhang mit EMB wurden bei allen NGS-Proben zwei SNPs mit 100 % alternativem Allel (*embB* L74R und *embB* D328G) detektiert. Bei der OFX-Resistenzbestimmung anhand der NGS-Daten traten deutliche Unterschiede zwischen den Proben auf. In den NGS-Daten der Probe 2 konnte die Mutation *gyrA* D94A bei etwa 99 % der ausgerichteten NGS-Sequenzfragmente detektiert werden. In den NGS-Daten der Probe 3 wurden zwei Mutationen im Verhältnis 6:4 beobachtet. Die Mutation *gyrA* D94A wurde von 40 % der ausgerichteten NGS-Sequenzfragmente vorhergesagt, die Mutation *gyrA* A90V konnte bei 60 % der ausgerichteten NGS-Sequenzfragmente detektiert werden. In den NGS-Daten der Probe 4 traten drei OFX-Resistenzmutationen parallel auf. Die Mutation *gyrA* D94A konnte bei etwa 5 % der ausgerichteten NGS-Sequenzfragmente beobachtet werden, die Mutation *gyrA* A90V bei 85 % der ausgerichteten NGS-Sequenzfragmente und neu dazu kam die Mutation *gyrA* D94N, die bei 10 % der ausgerichteten NGS-Sequenzfragmente beobachtet wurde. Ein Vergleich der Genomähnlichkeiten zeigte nah verwandte Stämme. Die NGS-Proben 2 und 3 zeigten zehn gemeinsame SNPs, die in keiner der anderen NGS-Proben detektiert wurden (Abbildung F.17).

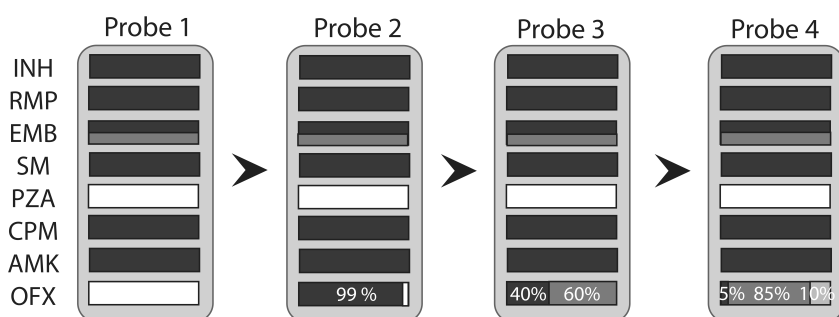


Abbildung F.16: Entwicklung des Resistenzprofils von Patient L anhand von NGS-Daten. Dunkelgrau = Resistenz-SNP 1; Grau = Resistenz-SNP 2; Hellgrau = Resistenz-SNP 3; Weiß = kein SNP

F.13 Patient M

Bei Patient M zeigte das phänotypische Resistenzprofil eine schrittweise Erweiterung. Ab Probe 2 waren die Proben neben INH-, RMP- und SM- auch EMB-

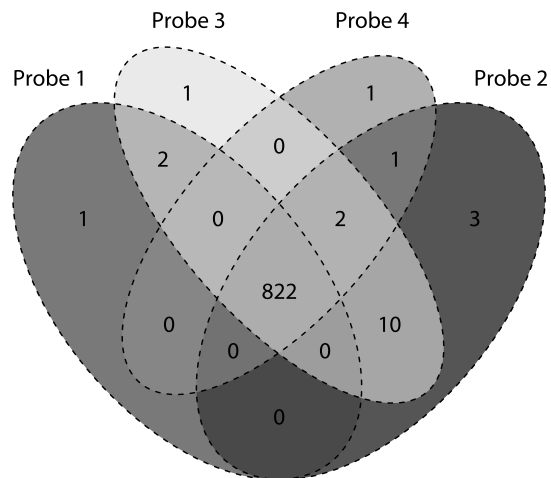


Abbildung F.17: Venn-Diagramm der SNP-Positionen der VCF-Dateien von Patient L. Alle NGS-Proben zeigten ungefähr dieselben SNPs. Die NGS-Probe 2 und 3 teilten zehn SNPs, die in keiner der anderen NGS-Proben detektiert wurden.

resistent. Ab Probe 3 kamen phänotypisch Resistenzen gegen PZA und OFX hinzu. In Probe 4 und 5 änderte sich das Resistenzprofil nicht mehr (Tabelle F.13).

Die Resistenzentwicklung anhand der NGS-Daten von Patient M ist in Abbildung F.18 dargestellt.

Anhand der NGS-Daten der Probe 1 konnten zwei INH-Resistenzmutationen detektiert werden. Die beiden Mutationen *katG* S315T und *fabG1* -15 C → T standen im Verhältnis 8:2. Ein ähnliches Verhältnis konnte bei den Mutationen *rpoB* S450L (75 %) und *rpoB* H445D beobachtet werden, die für die RMP-Resistenz verantwortlich waren. Außerdem konnte eine PZA-Resistenzmutation (*pncA* I133T) bei ungefähr 85 % der ausgerichteten NGS-Sequenzfragmente beobachtet werden. Weiterhin konnten anhand der NGS-Daten der Probe 1 die Mutationen *embB* M306V (EMB) und *rpsL* K43R (SM) bei 100 % der ausgerichteten NGS-Sequenzfragmente detektiert werden. Der Mischungsdetektor zeigte eine Mischung zweier Stämme der Linie Beijing an (Linie 2.2.1).

In den NGS-Daten der Probe 2 befanden sich die Hauptmutationen des Stammes aus Probe 1, während die niederfrequenten Mutationen nicht detektiert wurden.

Tabelle F.13: Entwicklung des phänotypischen Resistenzprofils der Proben von Patient M

Probe	INH	EMB	RMP	SM	PZA	OFX	CPM	AMK
1	R	S	R	R	S	S	S	S
2	R	R	R	R	S	S	S	S
3	R	R	R	R	R	R	S	S
4	R	R	R	R	R	R	S	S
5	R	R	R	R	R	R	S	S

Änderungen des Resistenzstatus in Grau. Abkürzungen: INH-Isoniazid, RMP-Rifampicin, EMB-Ethambutol, PZA-Pyrazinamid, SM-Streptomycin, OFX-Ofloxacin, CPM-Capreomycin, AMK-Amikacin, R-resistent, S-sensibel

Bei der NGS-Probe 3 wurde anhand der NGS-Daten eine zweite niederfrequente EMB resistenzvermittelnde Mutation detektiert. Drei Prozent der ausgerichteten NGS-Sequenzfragmente zeigten die Mutation *embB* D328G. Die PZA resistenzvermittelnde Mutation wurde von 94 % der ausgerichteten NGS-Sequenzfragmente angezeigt. Zusätzlich wurde eine neue Mutation *gyrA* A90V, die für eine OFX-Resistenz verantwortlich ist, bei 99 % der ausgerichteten NGS-Sequenzfragmente detektiert.

Anhand der NGS-Daten der Probe 4 konnte dasselbe Mutationsmuster wie bei der NGS-Probe 2 mit zusätzlichen Mutationen gegen OFX detektiert werden. Die Mutation *gyrA* A90V wurde von 85 % der ausgerichteten NGS-Sequenzfragmente angezeigt. Die Mutationen *gyrA* D94A und *gyrA* S91P wurden von 8 % bzw. 2 % angezeigt. Die NGS-Daten von Probe 5 ähnelte der von Probe 4 mit der Ausnahme, dass die Frequenzen der OFX-Mutationen variierten:

- *gyrA* A90V $\hat{=}$ 90 %
- *gyrA* D94A $\hat{=}$ 5 %
- *gyrA* S91P $\hat{=}$ 5 %

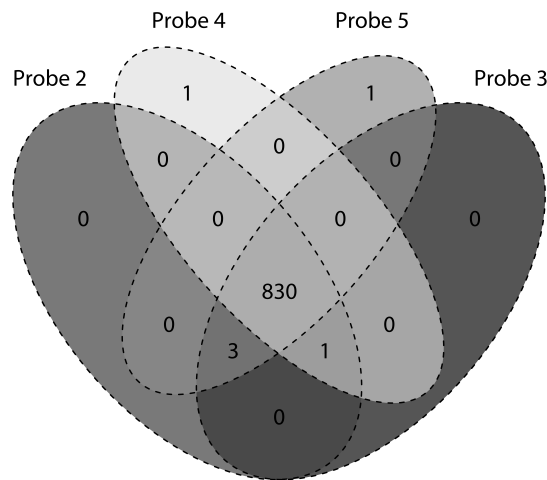


Abbildung F.19: Venn-Diagramm der SNP-Positionen der VCF-Dateien von Patient M. Alle aufgeführten NGS-Proben zeigten ungefähr dieselben SNPs.

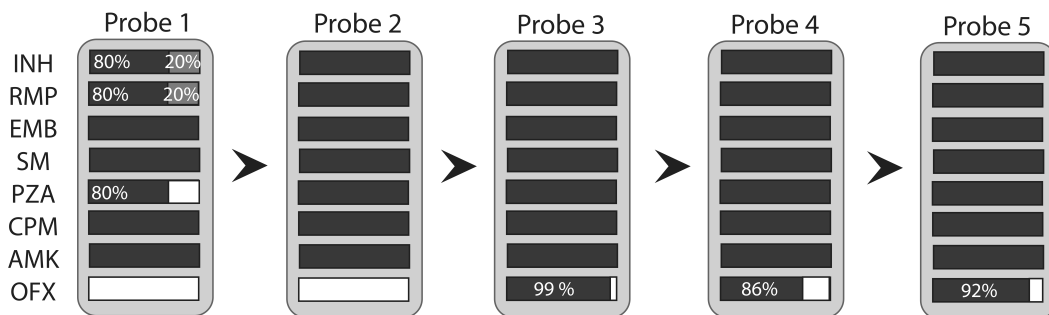


Abbildung F.18: Entwicklung des Resistenzprofils von Patient M anhand von NGS-Daten. Dunkelgrau = Resistenz-SNP 1; Grau = Resistenz-SNP 2; Weiß = kein SNP

**G Nachsequenzierte Proben mit nicht
eindeutigem MIRU-VNTR
Typisierungsmuster**

Tabelle G.1: Ergebnisse des Mischungsdetektors für die nachsequenzierten Proben mit nicht eindeutigem MIRU-VNTR Typisierungsergebnis.

ID	Population	MIRU	Mix	Clust1	Clust2	Clust3	Geno1	Freq	Geno2	Freq	Geno3	Freq	INH	RMP	FZA	SM	EMB	FQ	AMK, KAN, CPR
112-04	Turkmenistan	nicht ident	YES	0,996	0,793	0,205	Ural	80 %	Beijing	20 %			80 % und 100 %	80 %	20 %	80 % und 20 %	80 %		
11364-03	Nakus	nicht ident	NO	0,995			4.5	98 %					10 %	100 %	10 %	10 %	10 %		
1237-02	Turkmenistan	nicht ident	YES	0,948	0,122		Haarlem	100 %	Beijing	10 %			10 %	10 %	10 %	10 %	10 %		
2393-06	Nakus	Mix?	NO	0,989			Haarlem	90 %					100 %	100 %	10 %	100 %	100 %		
3674-04	Nakus	Mix	NO	0,991			Beijing	100 %					100 %	100 %	100 %	100 %	100 %		6 %
4247-09	Suessland	nicht ident	YES	0,938	0,128		4.4.1.1	90 %	Beijing	10 %			90 %	90 %	90 %	90 %	90 %		
4980-09	Suessland	nicht ident	YES	0,996	0,784	0,199	Beijing	80 %	LAM	20 %			100 %	100 %	20 %	20 %	100 %		
436-10	Suessland	nicht ident	NO	0,995			Beijing	100 %					100 %	100 %	100 %	100 %	100 %		100 %
4483-05	Nakus	nicht ident	NO	0,991			Beijing	100 %					100 %	100 %	100 %	100 %	100 %		
4541-04	Nakus	nicht ident	YES	0,992	0,482		Haarlem	55 %	Beijing	45 %			100 %	100 %	50 %	50 %	50 %		15 %, 38 %
4689-09	Suessland	nicht ident	YES	0,996	0,619	0,375	4.3.2.1/LAM	60 %	4.3.4.2.1/LAM	40 %			60 %	100 %	50 %	60 %	50 %		
4702-09	Suessland	nicht ident	YES	0,991	0,814	0,187	S-type	80 %	LAM	20 %			50 %	10 %	10 %	50 %	50 %		35 %
510-02	Turkmenistan	nicht ident	YES	0,995	0,507		LAM	50 %	Beijing	50 %			50 %	10 %, 10 %, 10 %	10 %	50 %	50 %		
5787-09	Suessland	nicht ident	NO	0,995			LAM	80 %					100 %	100 %	100 %	100 %	100 %		
6235-04	Nakus	nicht ident	NO	0,994			Beijing	100 %					100 %	100 %	100 %	100 %	100 %		
6258-09	Suessland	nicht ident	NO	0,994			LAM	100 %					100 %	100 %	100 %	100 %	100 %		
6333-05	Nakus	Mix	NO	0,990			Beijing	100 %					100 %	100 %	100 %	100 %	100 %		
6334-05	Nakus	nicht ident	NO	0,994			Beijing	100 %					100 %	100 %	100 %	100 %	100 %		100 % (nur KAN)
6687-04	Nakus	Mix	NO	0,983			Haarlem	96 %					100 %	100 %	100 %	100 %	100 %		
7188-09	Suessland	nicht ident	YES	0,915	0,133		Beijing	85 %	4.1.1.3	15 %			100 %	100 %	100 %	100 %	100 %		
7194-09	Suessland	nicht ident	NO	0,993			Beijing	100 %					100 %	100 %	100 %	100 %	100 %		
7350-05	Nakus	nicht ident	NO	0,990			Beijing	100 %					100 %	100 %	100 %	100 %	100 %		
780-02	Turkmenistan	nicht ident	YES	0,943	0,113		4.5	88 %	LAM	12 %			100 %	100 %	100 %	100 %	100 %		
8136-04	Nakus	Mix?	NO	0,993			4.8	99 %					100 %	100 %	100 %	100 %	100 %		
8405-01	Turkmenistan	nicht ident	YES	0,994	0,614	0,321	Haarlem	60 %	S-type	30 %	Beijing	10 %	80 %	10 %	10 %	10 %	10 %		
8444-05	Nakus	nicht ident	YES	0,991	0,798	0,196	Beijing	85 %	4.8	15 %			80 %, 20 %	100 %	85 %	85 %	100 %		20 %
8654-04	Nakus	Mix?	NO	0,990			4.8	98 %					100 %	100 %	100 %	100 %	100 %		
9548-01	Turkmenistan	nicht ident	YES	0,943	0,121		Ural	88 %	Beijing	12 %			100 %	12 %	12 %	12 %	100 %		
9826-01	Turkmenistan	nicht ident	NO	0,974			LAM	93 %	Beijing	7 %			100 %	100 %	100 %	100 %	100 %		
9964-05	Nakus	nicht ident	NO	0,991			Beijing	100 %					100 %	100 %	100 %	100 %	100 %		

Literaturverzeichnis

- [1] ABEL, H. J. ; DUNCAVAGE, E. : Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches. In: *Cancer genetics* 206 (2013), Dez., Nr. 12, 432–440. <http://dx.doi.org/10.1016/j.cancergen.2013.11.002>. – DOI 10.1016/j.cancergen.2013.11.002. – ISSN 2210–7762 **6.1**
- [2] ABYZOV, A. ; URBAN, A. E. ; SNYDER, M. ; GERSTEIN, M. : CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. In: *Genome Research* 21 (2011), Jun., Nr. 6, S. 974–984. <http://dx.doi.org/10.1101/gr.114876.110>. – DOI 10.1101/gr.114876.110. – ISSN 1549–5469 **6.1**
- [3] ACUNA-VILLAORDUNA, C. ; VASSALL, A. ; HENOSTROZA, G. ; SEAS, C. ; GUERRA, H. ; VASQUEZ, L. ; MORCILLO, N. ; SARAVIDA, J. ; O'BRIEN, R. ; PERKINS, M. D. ; CUNNINGHAM, J. ; LLANOS-ZAVALAGA, L. ; GOTUZZO, E. : Cost-effectiveness analysis of introduction of rapid, alternative methods to identify multidrug-resistant tuberculosis in middle-income countries. In: *Clin. Infect. Dis.* 47 (2008), Aug., Nr. 4, S. 487–495. <http://dx.doi.org/10.1086/590010>. – DOI 10.1086/590010. – ISSN 1537–6591. – PMID: 18636955 **1.6, 6.2.2**
- [4] ADEY, A. ; MORRISON, H. G. ; ASAN, n. ; XUN, X. ; KITZMAN, J. O. ; TURNER, E. H. ; STACKHOUSE, B. ; MACKENZIE, A. P. ; CARUCCIO, N. C. ; ZHANG, X. ; SHENDURE, J. : Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. In: *Genome Biology* 11 (2010), Nr. 12, S. R119. <http://dx.doi.org/10.1186/gb-2010-11-12-r119>. – DOI 10.1186/gb-2010-11-12-r119. – ISSN 1474–760X **2.1**
- [5] AHMAD, S. ; MOKADDAS, E. ; AL-MUTAIRI, N. ; ELDEEN, H. S. ; MOHAMMADI, S. : Discordance across Phenotypic and Molecular Methods

- for Drug Susceptibility Testing of Drug-Resistant Mycobacterium tuberculosis Isolates in a Low TB Incidence Country. In: *PLoS ONE* 11 (2016), Apr., Nr. 4. <http://dx.doi.org/10.1371/journal.pone.0153563>. – DOI 10.1371/journal.pone.0153563. – ISSN 1932–6203 1.6, 1.9
- [6] ALTMANN, A. ; WEBER, P. ; BADER, D. ; PREUSS, M. ; BINDER, E. B. ; MÜLLER-MYHSOK, B. : A beginners guide to SNP calling from high-throughput DNA-sequencing data. 131 (2012), Aug., Nr. 10, 1541–1554. <http://dx.doi.org/10.1007/s00439-012-1213-z>. – DOI 10.1007/s00439-012-1213-z. – ISSN 0340–6717, 1432–1203 2.3
- [7] ANDO, H. ; KONDO, Y. ; SUETAKE, T. ; TOYOTA, E. ; KATO, S. ; MORI, T. ; KIRIKAE, T. : Identification of katG Mutations Associated with High-Level Isoniazid Resistance in Mycobacterium tuberculosis. In: *Antimicrobial Agents and Chemotherapy* 54 (2010), Mai, Nr. 5, 1793–1799. <http://dx.doi.org/10.1128/AAC.01691-09>. – DOI 10.1128/AAC.01691-09. – ISSN 0066–4804 3.1, 4.3
- [8] ANDO, H. ; MIYOSHI-AKIYAMA, T. ; WATANABE, S. ; KIRIKAE, T. : A silent mutation in mabA confers isoniazid resistance on Mycobacterium tuberculosis. In: *Molecular Microbiology* 91 (2014), Febr., Nr. 3, 538–547. <http://dx.doi.org/10.1111/mmi.12476>. – DOI 10.1111/mmi.12476. – ISSN 1365–2958 3.1, 4.3
- [9] APPLIED MATHS: *BioNumerics Seven: a unique software platform*. <http://www.applied-maths.com/bionumerics>. Version: Jul. 2011 6.1
- [10] BAKER, M. : De novo genome assembly: what every biologist should know. In: *Nature Methods* 9 (2012), Apr., Nr. 4, 333–337. <http://dx.doi.org/10.1038/nmeth.1935>. – DOI 10.1038/nmeth.1935. – ISSN 1548–7091 1.8, 1.8
- [11] BANGE, F.-C. ; HAHN, H. ; KAUFMANN, S. H. E. ; ULRICHS: 41 Mykobakterien. In: SUERBAUM, S. (Hrsg.) ; BURCHARD, G.-D. (Hrsg.) ; SCHULZ, T. F. (Hrsg.): *Medizinische Mikrobiologie und Infektiologie*. 7. Springer, März 2012. – ISBN 978–3–642–24166–6, S. 341–355 1.4
- [12] BARRICK, J. E. ; LENSKI, R. E.: Genome-wide Mutational Diversity in an Evolving Population of Escherichia coli. 74 (2009), Sept., Nr. 0, 119–

129. <http://dx.doi.org/10.1101/sqb.2009.74.018>. – DOI 10.1101/sqb.2009.74.018. – ISSN 0091–7451, 1943–4456 **1.6**
- [13] BAYM, M. ; KRYAZHIMSKIY, S. ; LIEBERMAN, T. D. ; CHUNG, H. ; DESAI, M. M. ; KISHONY, R. : Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes. In: *PLOS ONE* 10 (2015), Mai, Nr. 5, e0128036. <http://dx.doi.org/10.1371/journal.pone.0128036>. – DOI 10.1371/journal.pone.0128036. – ISSN 1932–6203 **2.1**
- [14] BECK, T. F. ; MULLIKIN, J. C. ; PROGRAM, o. b. o. t. N. C. S. ; BIESECKER, L. G.: Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants. In: *Clinical Chemistry* 62 (2016), Apr., Nr. 4, 647–654. <http://dx.doi.org/10.1373/clinchem.2015.249623>. – DOI 10.1373/clinchem.2015.249623. – ISSN 0009–9147, 1530–8561 **6.1**
- [15] BERNOULLI, J. : *Ars conjectandi*. Impensis Thurnisiorum, fratrum, 1713. – Google-Books-ID: kD4PAAAAQAAJ **5.2.2**
- [16] BILLINGTON, O. J. ; MCHUGH, T. D. ; GILLESPIE, S. H.: Physiological Cost of Rifampin Resistance Induced In Vitro in Mycobacterium tuberculosis. In: *Antimicrobial Agents and Chemotherapy* 43 (1999), Jan., Nr. 8, 1866–1869. <http://aac.asm.org/content/43/8/1866>. – ISSN 0066–4804, 1098–6596 **6.2.2**
- [17] BLOUIN, Y. ; HAUCK, Y. ; SOLER, C. ; FABRE, M. ; VONG, R. ; DEHAN, C. ; CAZAJOUS, G. ; MASSOURE, P.-L. ; KRAEMER, P. ; JENKINS, A. ; GARNOTEL, E. ; POURCEL, C. ; VERGNAUD, G. : Significance of the Identification in the Horn of Africa of an Exceptionally Deep Branching Mycobacterium tuberculosis Clade. In: *PLOS ONE* 7 (2012), Dez., Nr. 12, e52841. <http://dx.doi.org/10.1371/journal.pone.0052841>. – DOI 10.1371/journal.pone.0052841. – ISSN 1932–6203 **1.5**
- [18] BRADLEY, P. ; GORDON, N. C. ; WALKER, T. M. ; DUNN, L. ; HEYS, S. ; HUANG, B. ; EARLE, S. ; PANKHURST, L. J. ; ANSON, L. ; CESARE, M. de ; PIAZZA, P. ; VOTINTSEVA, A. A. ; GOLUBCHIK, T. ; WILSON, D. J. ; WYLLIE, D. H. ; DIEL, R. ; NIEMANN, S. ; FEUERRIEGEL, S. ; KOHL, T. A. ; ISMAIL, N. ; OMAR, S. V. ; SMITH, E. G. ; BUCK, D. ; MCV EAN, G. ; WALKER, A. S. ; PETO, T. E. A. ; CROOK, D. W. ; IQBAL, Z. : Rapid

- antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. In: *Nature Communications* 6 (2015), Dez., 10063. <http://dx.doi.org/10.1038/ncomms10063>. – DOI 10.1038/ncomms10063 1.8, 3.3, 6.1
- [19] BREHMER, H. : *Die chronische Lungenschwindsucht und Tuberkulose der Lunge ihre Ursache und ihre Heilung*. 1857 1.1
- [20] BRENNAN, P. J. ; NIKAIDO, H. : The envelope of mycobacteria. In: *Annual Review of Biochemistry* 64 (1995), S. 29–63. <http://dx.doi.org/10.1146/annurev.bi.64.070195.000333>. – DOI 10.1146/annurev.bi.64.070195.000333. – ISSN 0066–4154 1.4
- [21] BROWN, A. C. ; BRYANT, J. M. ; EINER-JENSEN, K. ; HOLDSTOCK, J. ; HOUNIET, D. T. ; CHAN, J. Z. M. ; DEPLEDGE, D. P. ; NIKOLAYEVSKYY, V. ; BRODA, A. ; STONE, M. J. ; CHRISTIANSEN, M. T. ; WILLIAMS, R. ; MCANDREW, M. B. ; TUTILL, H. ; BROWN, J. ; MELZER, M. ; ROSMARIN, C. ; MCHUGH, T. D. ; SHORTEN, R. J. ; DROBNIIEWSKI, F. ; SPEIGHT, G. ; BREUER, J. : Rapid Whole-Genome Sequencing of *Mycobacterium tuberculosis* Isolates Directly from Clinical Samples. In: *Journal of Clinical Microbiology* 53 (2015), Jul., Nr. 7, S. 2230–2237. <http://dx.doi.org/10.1128/JCM.00486-15>. – DOI 10.1128/JCM.00486-15. – ISSN 1098–660X 1.8, 6.2.2
- [22] CAMINERO, J. A. ; PENA, M. J. ; CAMPOS-HERRERO, M. I. ; RODRÍGUEZ, J. C. ; AFONSO, O. ; MARTIN, C. ; PAVÓN, J. M. ; TORRES, M. J. ; BURGOS, M. ; CABRERA, P. ; SMALL, P. M. ; ENARSON, D. A.: Exogenous Reinfection with Tuberculosis on a European Island with a Moderate Incidence of Disease. In: *American Journal of Respiratory and Critical Care Medicine* 163 (2001), März, Nr. 3, 717–720. <http://dx.doi.org/10.1164/ajrccm.163.3.2003070>. – DOI 10.1164/ajrccm.163.3.2003070. – ISSN 1073–449X 1.7
- [23] CANETTI, G. ; FOX, W. ; KHOMENKO, A. ; MAHLER, H. T. ; MENON, N. K. ; MITCHISON, D. A. ; RIST, N. ; ŠMELEV, N. A.: Advances in techniques of testing mycobacterial drug sensitivity, and the use of sensitivity tests in tuberculosis control programmes. In: *Bulletin of the World Health*

- Organization* 41 (1969), Nr. 1, 21–43. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2427409/>. – ISSN 0042–9686 **6.2.2**
- [24] CASALI, N. ; NIKOLAYEVSKYY, V. ; BALABANOVA, Y. ; IGNATYEVA, O. ; KONTSEVAYA, I. ; HARRIS, S. R. ; BENTLEY, S. D. ; PARKHILL, J. ; NEJENTSEV, S. ; HOFFNER, S. E. ; HORSTMANN, R. D. ; BROWN, T. ; DROBNIIEWSKI, F. : Microevolution of extensively drug-resistant tuberculosis in Russia. In: *Genome Research* 22 (2012), Apr., Nr. 4, 735–745. <http://dx.doi.org/10.1101/gr.128678.111>. – DOI 10.1101/gr.128678.111. – ISSN 1088–9051 **5.1.3, B.2**
- [25] CASALI, N. ; NIKOLAYEVSKYY, V. ; BALABANOVA, Y. ; HARRIS, S. R. ; IGNATYEVA, O. ; KONTSEVAYA, I. ; CORANDER, J. ; BRYANT, J. ; PARKHILL, J. ; NEJENTSEV, S. ; HORSTMANN, R. D. ; BROWN, T. ; DROBNIIEWSKI, F. : Evolution and transmission of drug-resistant tuberculosis in a Russian population. In: *Nature Genetics* 46 (2014), März, Nr. 3, 279–286. <http://dx.doi.org/10.1038/ng.2878>. – DOI 10.1038/ng.2878. – ISSN 1061–4036 **4.3, 5.1.3, B.2**
- [26] CHAN, J. Z.-M. ; SERGEANT, M. J. ; LEE, O. Y.-C. ; MINNIKIN, D. E. ; BESRA, G. S. ; PAP, I. ; SPIGELMAN, M. ; DONOGHUE, H. D. ; PALLEN, M. J.: Metagenomic Analysis of Tuberculosis in a Mummy. In: *New England Journal of Medicine* 369 (2013), Jul., Nr. 3, 289–290. <http://dx.doi.org/10.1056/NEJMc1302295>. – DOI 10.1056/NEJMc1302295. – ISSN 0028–4793 **6.2.1**
- [27] CHRYSSANTHOU, E. ; ÄNGEBY, K. : The GenoType®MTBDRplus assay for detection of drug resistance in Mycobacterium tuberculosis in Sweden. In: *APMIS* 120 (2012), Mai, Nr. 5, 405–409. <http://dx.doi.org/10.1111/j.1600-0463.2011.02845.x>. – DOI 10.1111/j.1600–0463.2011.02845.x. – ISSN 1600–0463 **3.3**
- [28] COCK, P. J. A. ; FIELDS, C. J. ; GOTO, N. ; HEUER, M. L. ; RICE, P. M.: The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. In: *Nucleic Acids Res* 38 (2010), Apr., Nr. 6, 1767–1771. <http://dx.doi.org/10.1093/nar/gkp1137>. – DOI 10.1093/nar/gkp1137. – ISSN 0305–1048 **2.2.1**

- [29] COHEN, T. ; HELDEN, P. D. ; WILSON, D. ; COLIJN, C. ; McLAUGHLIN, M. M. ; ABUBAKAR, I. ; WARREN, R. M.: Mixed-strain mycobacterium tuberculosis infections and the implications for tuberculosis treatment and control. In: *Clinical Microbiology Reviews* 25 (2012), Okt., Nr. 4, S. 708–719. <http://dx.doi.org/10.1128/CMR.00021-12>. – DOI 10.1128/CMR.00021-12. – ISSN 1098-6618 1.7, 1.7, 1.4, 6.2.1
- [30] COHEN, T. ; WILSON, D. ; WALLENGREN, K. ; SAMUEL, E. Y. ; MURRAY, M. : Mixed-Strain Mycobacterium tuberculosis Infections among Patients Dying in a Hospital in KwaZulu-Natal, South Africa. In: *Journal of Clinical Microbiology* 49 (2011), Jan., Nr. 1, 385–388. <http://dx.doi.org/10.1128/JCM.01378-10>. – DOI 10.1128/JCM.01378-10. – ISSN 0095-1137, 1098-660X 1.7, 1.7
- [31] COLDITZ GA ; BREWER TF ; BERKEY CS ; ET AL: Efficacy of bcg vaccine in the prevention of tuberculosis: Meta-analysis of the published literature. In: *JAMA* 271 (1994), März, Nr. 9, 698–702. <http://dx.doi.org/10.1001/jama.1994.03510330076038>. – DOI 10.1001/jama.1994.03510330076038. – ISSN 0098-7484 1.1
- [32] COLL, F. ; McNERNEY, R. ; GUERRA-ASSUNÇÃO, J. A. ; GLYNN, J. R. ; PERDIGÃO, J. a. ; VIVEIROS, M. ; PORTUGAL, I. ; PAIN, A. ; MARTIN, N. ; CLARK, T. G.: A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. In: *Nature Communications* 5 (2014), Sept. <http://dx.doi.org/10.1038/ncomms5812>. – DOI 10.1038/ncomms5812 1.5, 3.4, 4.2, 5.1.4, 6.1
- [33] COLL, F. ; McNERNEY, R. ; PRESTON, M. D. ; GUERRA-ASSUNÇÃO, J. A. ; WARRY, A. ; HILL-CAWTHORNE, G. ; MALLARD, K. ; NAIR, M. ; MIRANDA, A. ; ALVES, A. ; PERDIGÃO, J. a. ; VIVEIROS, M. ; PORTUGAL, I. ; HASAN, Z. ; HASAN, R. ; GLYNN, J. R. ; MARTIN, N. ; PAIN, A. ; CLARK, T. G.: Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. In: *Genome Medicine* 7 (2015), Mai, Nr. 1, 51. <http://dx.doi.org/10.1186/s13073-015-0164-0>. – DOI 10.1186/s13073-015-0164-0. – ISSN 1756-994X 1.8, 1.8, 3.5, 4.5, 6.1
- [34] COMAS, I. n. ; HOMOLKA, S. ; NIEMANN, S. ; GAGNEUX, S. : Genotyping of Genetically Monomorphic Bacteria: DNA Sequencing in Mycobacteri-

- um tuberculosis Highlights the Limitations of Current Methodologies. In: *PLoS ONE* 4 (2009), Nov., Nr. 11, e7815. <http://dx.doi.org/10.1371/journal.pone.0007815>. – DOI 10.1371/journal.pone.0007815 4.2
- [35] COURY, C. ; CONSTANS, P. ; PARROT, R. ; BARON, A. : Results of treatment with rifampicin-isoniazid in 100 tubercular patients with sensitive bacilli a 6 to 24 month controlled experiment. In: *Revue De Tuberculose Et De Pneumologie* 34 (1970), Aug., Nr. 5, S. 688–697. – ISSN 0035–1792 1.1
- [36] COX, H. S. ; SIBILIA, C. ; FEUERRIEGEL, S. ; KALON, S. ; POLONSKY, J. ; KHAMRAEV, A. K. ; RÜSCH-GERDES, S. ; MILLS, C. ; NIEMANN, S. : Emergence of Extensive Drug Resistance during Treatment for Multidrug-Resistant Tuberculosis. In: *New England Journal of Medicine* 359 (2008), Nov., Nr. 22, 2398–2400. <http://dx.doi.org/10.1056/NEJMc0805644>. – DOI 10.1056/NEJMc0805644. – ISSN 0028–4793 5.2.5, 5.10, 6.2.2
- [37] DANECEK, P. ; AUTON, A. ; ABECASIS, G. ; ALBERS, C. A. ; BANKS, E. ; DEPRISTO, M. A. ; HANDSAKER, R. E. ; LUNTER, G. ; MARTH, G. T. ; SHERRY, S. T. ; MCVEAN, G. ; DURBIN, R. : The variant call format and VCFtools. In: *Bioinformatics* 27 (2011), Aug., Nr. 15, 2156–2158. <http://dx.doi.org/10.1093/bioinformatics/btr330>. – DOI 10.1093/bioinformatics/btr330. – ISSN 1367–4803. – PMID: 21653522 PMCID: PMC3137218 2.2.3
- [38] DATABANK, T. W.: *WDI Database Archives (beta)| World DataBank*. <http://databank.worldbank.org/data/reports.aspx?source=WDI-Archives> B.1
- [39] DAVIES, P. D.: Tuberculosis: the global epidemic. In: *Journal of the Indian Medical Association* 98 (2000), März, Nr. 3, S. 100–102. – ISSN 0019–5847 1.1
- [40] DELCHER, A. L. ; KASIF, S. ; FLEISCHMANN, R. D. ; PETERSON, J. ; WHITE, O. ; SALZBERG, S. L.: Alignment of whole genomes. In: *Nucleic Acids Research* 27 (1999), Jan., Nr. 11, 2369–2376. <http://dx.doi.org/10.1093/nar/27.11.2369>. – DOI 10.1093/nar/27.11.2369. – ISSN 0305–1048, 1362–4962

- [41] DELCHER, A. L. ; PHILLIPPY, A. ; CARLTON, J. ; SALZBERG, S. L.: Fast algorithms for large-scale genome alignment and comparison. In: *Nucleic Acids Research* 30 (2002), Jun., Nr. 11, S. 2478–2483. – ISSN 1362–4962 **3.1**
- [42] DEPRISTO, M. A. ; BANKS, E. ; POPLIN, R. ; GARIMELLA, K. V. ; MAGUIRE, J. R. ; HARTL, C. ; PHILIPPAKIS, A. A. ; ANGEL, G. del ; RIVAS, M. A. ; HANNA, M. ; MCKENNA, A. ; FENNELL, T. J. ; KERNYTSKY, A. M. ; SIVACHENKO, A. Y. ; CIBULSKIS, K. ; GABRIEL, S. B. ; ALTSHULER, D. ; DALY, M. J.: A framework for variation discovery and genotyping using next-generation DNA sequencing data. In: *Nature genetics* 43 (2011), Mai, Nr. 5, S. 491–498. <http://dx.doi.org/10.1038/ng.806>. – DOI 10.1038/ng.806. – ISSN 1546–1718 **2.3**
- [43] DEUTSCHE GESELLSCHAFT FÜR DIE VEREINTEN NATIONEN (Hrsg.) ; BWV BERLINER WISSENSCHAFTS-VERLAG GMBH (Hrsg.): *Arbeit und menschliche Entwicklung*. Deutsche Ausgabe. Berlin : Berliner Wissenschafts-Verlag GmbH, 2015 (Bericht über die menschliche Entwicklung 2015). – ISBN 978–3–8305–3618–5. – OCLC: 934153060 **2.2, B.1**
- [44] DEYCKE, G. : *Praktisches Lehrbuch der Tuberkulose*. Springer Berlin Heidelberg <http://link.springer.com/10.1007/978-3-642-99542-2>. – ISBN 978–3–642–98727–4 978–3–642–99542–2 **1.3**
- [45] DIEHL, R. : *Behandlungskosten für Tuberkulose und MDR-/XDR-Tuberkulose in Deutschland*. März 2014 **1.6**
- [46] DOMINIC J. REPICI: *CSV Comma Separated Value File Format - How To - Creativyst - Explored, Designed, Delivered. (sm)*. <http://www.creativyst.com/Doc/Articles/CSV/CSV01.htm>. Version: 2004 **2.2.4**
- [47] DONALD, P. R. ; HELDEN, P. D. v.: *Antituberculosis Chemotherapy*. Karger Medical and Scientific Publishers, 2011. – ISBN 978–3–8055–9628–2. – Google-Books-ID: aZc7AQAAQBAJ **6.2.1**
- [48] ECMA INTERNATIONAL: *Standard ECMA-404: The JSON Data Interchange Format*. <https://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf>. Version: Okt. 2013 **2.2.5**

- [49] EDOCEO, INC.: *Comma Separated Values (CSV) Standard File Format*. <http://edoceo.com/utilitas/csv-file-format>. Version: 2004
- [50] EMBDEN, J. D. ; CAVE, M. D. ; CRAWFORD, J. T. ; DALE, J. W. ; EISENACH, K. D. ; GICQUEL, B. ; HERMANS, P. ; MARTIN, C. ; McADAM, R. ; SHINNICK, T. M.: Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. In: *Journal of Clinical Microbiology* 31 (1993), Febr., Nr. 2, S. 406–409. – ISSN 0095–1137 **1.5**
- [51] EMBL-EBI: *European Nucleotide Archive < EMBL-EBI*. <http://www.ebi.ac.uk/ena> **2.4**
- [52] ESTER, M. ; KRIEGEL, H.-P. ; SANDER, J. ; XU, X. : A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. (1996). <https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf> **5.1.1, 5.1.1, 5.1.2**
- [53] EWING, B. ; GREEN, P. : Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. In: *Genome Res.* 8 (1998), Jan., Nr. 3, 186–194. <http://dx.doi.org/10.1101/gr.8.3.186>. – DOI 10.1101/gr.8.3.186. – ISSN 1088–9051, 1549–5469. – PMID: 9521922 **2.2.1**
- [54] EYRE, D. W. ; CULE, M. L. ; GRIFFITHS, D. ; CROOK, D. W. ; PETO, T. E. A. ; WALKER, A. S. ; WILSON, D. J.: Detection of Mixed Infection from Bacterial Whole Genome Sequence Data Allows Assessment of Its Role in Clostridium difficile Transmission. In: *PLOS Computational Biology* 9 (2013), Mai, Nr. 5, e1003059. <http://dx.doi.org/10.1371/journal.pcbi.1003059>. – DOI 10.1371/journal.pcbi.1003059. – ISSN 1553–7358 **6.2.1**
- [55] FEUERRIEGEL, S. ; OBERHAUSER, B. ; GEORGE, A. G. ; DAFAR, F. ; RICHTER, E. ; RÜSCH-GERDES, S. ; NIEMANN, S. : Sequence analysis for detection of first-line drug resistance in Mycobacterium tuberculosis strains from a high-incidence setting. In: *BMC Microbiology* 12 (2012), 90. <http://dx.doi.org/10.1186/1471-2180-12-90>. – DOI 10.1186/1471-2180-12-90. – ISSN 1471–2180 **4, 4.3, 4.3, 6.1, B.1**

- [56] FEUERRIEGEL, S. ; SCHLEUSENER, V. ; BECKERT, P. ; KOHL, T. A. ; MIOTTO, P. ; CIRILLO, D. M. ; CABIBBE, A. M. ; NIEMANN, S. ; FELLEBERG, K. : PhyResSE: a Web Tool Delineating Mycobacterium tuberculosis Antibiotic Resistance and Lineage from Whole-Genome Sequencing Data. In: *Journal of Clinical Microbiology* 53 (2015), Jun., Nr. 6, S. 1908–1914. <http://dx.doi.org/10.1128/JCM.00025-15>. – DOI 10.1128/JCM.00025-15. – ISSN 1098-660X 1.8, 3.4, 4.1, 4.4
- [57] FLAHERTY, P. ; NATSOULIS, G. ; MURALIDHARAN, O. ; WINTERS, M. ; BUENROSTRO, J. ; BELL, J. ; BROWN, S. ; HOLODNIY, M. ; ZHANG, N. ; JI, H. P.: Ultrasensitive detection of rare mutations using next-generation targeted resequencing. In: *Nucleic Acids Research* 40 (2012), Jan., Nr. 1, S. e2. <http://dx.doi.org/10.1093/nar/gkr861>. – DOI 10.1093/nar/gkr861. – ISSN 1362-4962 6.2.2
- [58] FLUSBERG, B. A. ; WEBSTER, D. ; LEE, J. ; TRAVERS, K. ; OLIVARES, E. ; CLARK, T. A. ; KORLACH, J. ; TURNER, S. W.: Direct detection of DNA methylation during single-molecule, real-time sequencing. In: *Nature methods* 7 (2010), Jun., Nr. 6, 461–465. <http://dx.doi.org/10.1038/nmeth.1459>. – DOI 10.1038/nmeth.1459. – ISSN 1548-7091 1.8
- [59] FOLKVARDBSEN, D. B. ; SVENSSON, E. ; THOMSEN, V. .. ; RASMUSSEN, E. M. ; BANG, D. ; WERNGREN, J. ; HOFFNER, S. ; HILLEMANN, D. ; RIGOUTS, L. : Can Molecular Methods Detect 1% Isoniazid Resistance in Mycobacterium tuberculosis? In: *Journal of Clinical Microbiology* 51 (2013), Mai, Nr. 5, 1596–1599. <http://dx.doi.org/10.1128/JCM.00472-13>. – DOI 10.1128/JCM.00472-13. – ISSN 0095-1137 1.7
- [60] FOLKVARDBSEN, D. B. ; THOMSEN, V. .. ; RIGOUTS, L. ; RASMUSSEN, E. M. ; BANG, D. ; BERNAERTS, G. ; WERNGREN, J. ; TORO, J. C. ; HOFFNER, S. ; HILLEMANN, D. ; SVENSSON, E. : Rifampin Heteroresistance in Mycobacterium tuberculosis Cultures as Detected by Phenotypic and Genotypic Drug Susceptibility Test Methods. In: *Journal of Clinical Microbiology* 51 (2013), Dez., Nr. 12, 4220–4222. <http://dx.doi.org/10.1128/JCM.01602-13>. – DOI 10.1128/JCM.01602-13. – ISSN 0095-1137 1.7
- [61] FORD, C. B. ; SHAH, R. R. ; MAEDA, M. K. ; GAGNEUX, S. ; MURRAY, M. B. ; COHEN, T. ; JOHNSTON, J. C. ; GARDY, J. ; LIPSITCH, M. ;

- FORTUNE, S. M.: Mycobacterium tuberculosis mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. In: *Nat Genet* 45 (2013), Jul., Nr. 7, 784–790. <http://dx.doi.org/10.1038/ng.2656>. – DOI 10.1038/ng.2656. – ISSN 1061–4036 **1.5**
- [62] FRANÇA, L. T. C. ; CARRILHO, E. ; KIST, T. B. L.: A review of DNA sequencing techniques. In: *Quarterly Reviews of Biophysics* 35 (2002), Mai, Nr. 02. <http://dx.doi.org/10.1017/S0033583502003797>. – DOI 10.1017/S0033583502003797. – ISSN 0033–5835, 1469–8994 **1.8**
- [63] FROTHINGHAM, R. ; MEEKER-O’CONNELL, W. A.: Genetic diversity in the Mycobacterium tuberculosis complex based on variable numbers of tandem DNA repeats. In: *Microbiology (Reading, England)* 144 (Pt 5) (1998), Mai, S. 1189–1196. <http://dx.doi.org/10.1099/00221287-144-5-1189>. – DOI 10.1099/00221287-144-5-1189. – ISSN 1350–0872 **1.5**
- [64] GAGNEUX, S. : Fitness cost of drug resistance in Mycobacterium tuberculosis. In: *Clinical Microbiology and Infection* 15, Supplement 1 (2009), Jan., 66–68. <http://dx.doi.org/10.1111/j.1469-0691.2008.02685.x>. – DOI 10.1111/j.1469-0691.2008.02685.x. – ISSN 1198–743X **6.2.2**
- [65] GAGNEUX, S. ; DERIEMER, K. ; VAN, T. ; KATO-MAEDA, M. ; JONG, B. C. ; NARAYANAN, S. ; NICOL, M. ; NIEMANN, S. ; KREMER, K. ; GUTIERREZ, M. C. ; HILTY, M. ; HOPEWELL, P. C. ; SMALL, P. M.: Variable host-pathogen compatibility in Mycobacterium tuberculosis. In: *Proceedings of the National Academy of Sciences of the United States of America* 103 (2006), Febr., Nr. 8, 2869–2873. <http://dx.doi.org/10.1073/pnas.0511240103>. – DOI 10.1073/pnas.0511240103. – ISSN 0027–8424 **1.5**
- [66] GAN, M. ; LIU, Q. ; YANG, C. ; GAO, Q. ; LUO, T. : Deep Whole-Genome Sequencing to Detect Mixed Infection of Mycobacterium tuberculosis. In: *PLoS ONE* 11 (2016), Jul., Nr. 7. <http://dx.doi.org/10.1371/journal.pone.0159029>. – DOI 10.1371/journal.pone.0159029. – ISSN 1932–6203 **6.2.1**
- [67] GEORGII, H.-O. : *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Walter de Gruyter GmbH & Co KG, 2015. – ISBN 978–3–11–038686–8. – Google-Books-ID: 1pAFCgAAQBAJ **5.2.2, 5.2.2**

- [68] GERSTUNG, M. ; BEISEL, C. ; RECHSTEINER, M. ; WILD, P. ; SCHRAML, P. ; MOCH, H. ; BEERENWINKEL, N. : Reliable detection of subclonal single-nucleotide variants in tumour cell populations. In: *Nature Communications* 3 (2012), Mai, S. 811. <http://dx.doi.org/10.1038/ncomms1814>. – DOI 10.1038/ncomms1814. – ISSN 2041–1723 6.2.2
- [69] GLENN, T. C.: Field guide to next-generation DNA sequencers. In: *Molecular Ecology Resources* 11 (2011), Sept., Nr. 5, 759–769. <http://dx.doi.org/10.1111/j.1755-0998.2011.03024.x>. – DOI 10.1111/j.1755-0998.2011.03024.x. – ISSN 1755–0998 5.2.3
- [70] GRANGE, J. M. ; ZUMLA, A. : The global emergency of tuberculosis: what is the cause? In: *The Journal of the Royal Society for the Promotion of Health* 122 (2002), Jun., Nr. 2, S. 78–81. – ISSN 1466–4240 1.1
- [71] GRUNDMANN, K. ; FRIEDRICH, C. : Zur Geschichte der Tuberkulosebehandlung. In: *Pharmazie in unserer Zeit* 41 (2012), Jan., Nr. 1, 10–18. <http://dx.doi.org/10.1002/pauz.201100448>. – DOI 10.1002/pauz.201100448. – ISSN 1615–1003 1.1
- [72] GYGLI, S. M. ; BORRELL, S. ; TRAUNER, A. ; GAGNEUX, S. : Antimicrobial resistance in *Mycobacterium tuberculosis*: mechanistic and evolutionary perspectives. In: *FEMS microbiology reviews* (2017), März. <http://dx.doi.org/10.1093/femsre/fux011>. – DOI 10.1093/femsre/fux011. – ISSN 1574–6976 1.5
- [73] HADFIELD, J. : *CoreGenomics: Battle of the benchtops: MiSeq vs Ion vs 454 and bacterial genomics*. <http://core-genomics.blogspot.com/2012/04/battle-of-benchtopmiseq-vs-ion-vs-454.html>. Version: 2012 1.2
- [74] HANEKOM, M. ; STREICHER, E. M. ; BERG, D. Van d. ; COX, H. ; MC-DELMID, C. ; BOSMAN, M. ; PITTIUS, N. C. v. ; VICTOR, T. C. ; KIDD, M. ; SOOLINGEN, D. van ; HELDEN, P. D. ; WARREN, R. M.: Population Structure of Mixed *Mycobacterium tuberculosis* Infection Is Strain Genotype and Culture Medium Dependent. In: *PLoS ONE* 8 (2013), Jul., Nr. 7, e70178. <http://dx.doi.org/10.1371/journal.pone.0070178>. – DOI 10.1371/journal.pone.0070178 1.7, 6.2.1, 6.3

- [75] HARISMENDY, O. ; SCHWAB, R. B. ; BAO, L. ; OLSON, J. ; ROZENZHAK, S. ; KOTSOPOULOS, S. K. ; POND, S. ; CRAIN, B. ; CHEE, M. S. ; MESSER, K. ; LINK, D. R. ; FRAZER, K. A.: Detection of low prevalence somatic mutations in solid tumors with ultra-deep targeted sequencing. In: *Genome Biology* 12 (2011), Dez., Nr. 12, S. R124. <http://dx.doi.org/10.1186/gb-2011-12-12-r124>. – DOI 10.1186/gb-2011-12-12-r124. – ISSN 1474-760X 6.2.2
- [76] HEATHER, J. M. ; CHAIN, B. : The sequence of sequencers: The history of sequencing DNA. In: *Genomics* 107 (2016), Jan., Nr. 1, 1–8. <http://dx.doi.org/10.1016/j.ygeno.2015.11.003>. – DOI 10.1016/j.ygeno.2015.11.003. – ISSN 0888-7543 1.8
- [77] HERSHBERG, R. ; LIPATOV, M. ; SMALL, P. M. ; SHEFFER, H. ; NIEMANN, S. ; HOMOLKA, S. ; ROACH, J. C. ; KREMER, K. ; PETROV, D. A. ; FELDMAN, M. W. ; GAGNEUX, S. : High Functional Diversity in Mycobacterium tuberculosis Driven by Genetic Drift and Human Demography. In: *PLoS Biol* 6 (2008), Dez., Nr. 12, e311. <http://dx.doi.org/10.1371/journal.pbio.0060311>. – DOI 10.1371/journal.pbio.0060311 1.5, 1.7
- [78] HERSHKOVITZ, I. ; DONOGHUE, H. D. ; MINNIKIN, D. E. ; BESRA, G. S. ; LEE, O. Y.-C. ; GERNAEY, A. M. ; GALILI, E. ; ESHED, V. ; GREENBLATT, C. L. ; LEMMA, E. ; BAR-GAL, G. K. ; SPIGELMAN, M. : Detection and Molecular Characterization of 9000-Year-Old Mycobacterium tuberculosis from a Neolithic Settlement in the Eastern Mediterranean. In: *PLOS ONE* 3 (2008), Okt., Nr. 10, e3426. <http://dx.doi.org/10.1371/journal.pone.0003426>. – DOI 10.1371/journal.pone.0003426. – ISSN 1932-6203 1.1
- [79] HIATT, J. B. ; PRITCHARD, C. C. ; SALIPANTE, S. J. ; O'ROAK, B. J. ; SHENDURE, J. : Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. In: *Genome Research* 23 (2013), Mai, Nr. 5, S. 843–854. <http://dx.doi.org/10.1101/gr.147686.112>. – DOI 10.1101/gr.147686.112. – ISSN 1549-5469 6.2.2
- [80] HINGLEY-WILSON, S. M. ; CASEY, R. ; CONNELL, D. ; BREMANG, S. ; EVANS, J. T. ; HAWKEY, P. M. ; SMITH, G. E. ; JEPSON, A. ; PHILIP, S. ; KON, O. M. ; LALVANI, A. : Undetected Multidrug-Resistant Tuberculosis Amplified by First-line Therapy in Mixed Infection. In: *Emerging Infectious*

- Diseases* 19 (2013), Jul., Nr. 7, 1138–1141. <http://dx.doi.org/10.3201/eid1907.130313>. – DOI 10.3201/eid1907.130313. – ISSN 1080–6040, 1080–6059 1.7
- [81] HOMOLKA, S. ; POST, E. ; OBERHAUSER, B. ; GEORGE, A. G. ; WESTMAN, L. ; DAFAR, F. ; RÜSCH-GERDES, S. ; NIEMANN, S. : High genetic diversity among Mycobacterium tuberculosis complex strains from Sierra Leone. In: *BMC Microbiology* 8 (2008), Jun., Nr. 1, 103. <http://dx.doi.org/10.1186/1471-2180-8-103>. – DOI 10.1186/1471-2180-8-103. – ISSN 1471–2180 4, B.1
- [82] HOMOLKA, S. ; PROJAHN, M. ; FEUERRIEGEL, S. ; UBBEN, T. ; DIEL, R. ; NÜBEL, U. ; NIEMANN, S. : High Resolution Discrimination of Clinical Mycobacterium tuberculosis Complex Strains Based on Single Nucleotide Polymorphisms. In: *PLoS ONE* 7 (2012), Jul., Nr. 7, e39855. <http://dx.doi.org/10.1371/journal.pone.0039855>. – DOI 10.1371/journal.pone.0039855 1.5, 1.5, 3.4, 4.2, 4.4
- [83] HORNE, D. J. ; PINTO, L. M. ; ARENTZ, M. ; LIN, S.-Y. G. ; DESMOND, E. ; FLORES, L. L. ; STEINGART, K. R. ; MINION, J. : Diagnostic Accuracy and Reproducibility of WHO-Endorsed Phenotypic Drug Susceptibility Testing Methods for First-Line and Second-Line Antituberculosis Drugs. In: *Journal of Clinical Microbiology* 51 (2013), Febr., Nr. 2, 393–401. <http://dx.doi.org/10.1128/JCM.02724-12>. – DOI 10.1128/JCM.02724-12. – ISSN 0095–1137 1.6
- [84] HUYEN, M. N. T. ; KREMER, K. ; LAN, N. T. N. ; COBELENS, F. G. J. ; BUU, T. N. ; DUNG, N. H. ; CAWS, M. ; TIEMERSMA, E. W. ; SOOLINGEN, D. v.: Mixed Tuberculosis Infections in Rural South Vietnam. In: *Journal of Clinical Microbiology* 50 (2012), Jan., Nr. 5, 1586–1592. <http://dx.doi.org/10.1128/JCM.00434-12>. – DOI 10.1128/JCM.00434-12. – ISSN 0095–1137, 1098–660X 1.7
- [85] ILLUMINA, I. : *Paired-End Sequencing*. https://www.illumina.com/technology/next-generation-sequencing/paired-end-sequencing_assay.html. Version: 2017 1.8
- [86] IWAI, H. ; KATO-MIYAZAWA, M. ; KIRIKAE, T. ; MIYOSHI-AKIYAMA, T. : CASTB (the comprehensive analysis server for the Mycobacterium tu-

- berculosis complex): A publicly accessible web server for epidemiological analyses, drug-resistance prediction and phylogenetic comparison of clinical isolates. In: *Tuberculosis (Edinburgh, Scotland)* 95 (2015), Dez., Nr. 6, S. 843–844. <http://dx.doi.org/10.1016/j.tube.2015.09.002>. – DOI 10.1016/j.tube.2015.09.002. – ISSN 1873–281X **1.8, 3.1**
- [87] JAGIELSKI, T. ; BAKUŁA, Z. ; ROESKE, K. ; KAMIŃSKI, M. ; NAPIÓRKOWSKA, A. ; AUGUSTYNOWICZ-KOPEĆ, E. ; ZWOLSKA, Z. ; BIELECKI, J. : Detection of mutations associated with isoniazid resistance in multidrug-resistant *Mycobacterium tuberculosis* clinical isolates. In: *Journal of Antimicrobial Chemotherapy* (2014), Mai, dku161. <http://dx.doi.org/10.1093/jac/dku161>. – DOI 10.1093/jac/dku161. – ISSN 0305–7453, 1460–2091 **6.1**
- [88] JANITZ, M. (Hrsg.): *Next-generation genome sequencing: towards personalized medicine*. 1. Reprint. Weinheim : Wiley-Blackwell Wiley-VCH, 2010. – ISBN 978–3–527–32090–5. – OCLC: 767958127 **6.1**
- [89] JARLIER, V. ; NIKAIDO, H. : Mycobacterial cell wall: structure and role in natural resistance to antibiotics. In: *FEMS microbiology letters* 123 (1994), Okt., Nr. 1-2, S. 11–18. – ISSN 0378–1097 **1.4**
- [90] JASMER, R. M. ; BOZEMAN, L. ; SCHWARTZMAN, K. ; CAVE, M. D. ; SAUKKONEN, J. J. ; METCHOCK, B. ; KHAN, A. ; BURMAN, W. J. ; TUBERCULOSIS TRIALS CONSORTIUM: Recurrent tuberculosis in the United States and Canada: relapse or reinfection? In: *American Journal of Respiratory and Critical Care Medicine* 170 (2004), Dez., Nr. 12, S. 1360–1366. <http://dx.doi.org/10.1164/rccm.200408-10810C>. – DOI 10.1164/rccm.200408-10810C. – ISSN 1073–449X **6.2.1**
- [91] JENTGENS, H. : *Lungentuberkulose*. Springer-Verlag, 2013. – ISBN 978–3–642–67664–2. – Google-Books-ID: 0NynBgAAQBAJ **6.2.2**
- [92] JIANG, Y. ; WANG, Y. ; BRUDNO, M. : PRISM: pair-read informed split-read mapping for base-pair level detection of insertion, deletion and structural variants. In: *Bioinformatics (Oxford, England)* 28 (2012), Okt., Nr. 20, S. 2576–2583. <http://dx.doi.org/10.1093/bioinformatics/bts484>. – DOI 10.1093/bioinformatics/bts484. – ISSN 1367–4811 **6.1**

- [93] JOLLEY, K. A. ; MAIDEN, M. C.: BIGSdb: Scalable analysis of bacterial genome variation at the population level. In: *BMC Bioinformatics* 11 (2010), Dez., 595. <http://dx.doi.org/10.1186/1471-2105-11-595>. – DOI 10.1186/1471-2105-11-595. – ISSN 1471-2105 **3.1**
- [94] JR, A. M. D.: Immunopathogenesis of Pulmonary Tuberculosis. In: *Hospital Practice* 28 (1993), Jan., Nr. 1, 51–58. <http://dx.doi.org/10.1080/21548331.1993.11442738>. – DOI 10.1080/21548331.1993.11442738. – ISSN 2154-8331 **1.3**
- [95] KAMERBEEK, J. ; SCHOULS, L. ; KOLK, A. ; AGTERVELD, M. van ; SOOLINGEN, D. van ; KUIJPER, S. ; BUNSCHOTEN, A. ; MOLHUIZEN, H. ; SHAW, R. ; GOYAL, M. ; EMBDEN, J. van: Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. In: *Journal of Clinical Microbiology* 35 (1997), Apr., Nr. 4, S. 907–914. – ISSN 0095-1137 **1.5**
- [96] KAPPELMAN, J. ; ALÇIÇEK, M. C. ; KAZANCI, N. ; SCHULTZ, M. ; ÖZKUL, M. ; ŞEN, c. : First *Homo erectus* from Turkey and implications for migrations into temperate Eurasia. In: *American Journal of Physical Anthropology* 135 (2008), Jan., Nr. 1, 110–116. <http://dx.doi.org/10.1002/ajpa.20739>. – DOI 10.1002/ajpa.20739. – ISSN 1096-8644 **1.1**
- [97] KAUFMANN, S. H. E. (Hrsg.) ; HAHN, H. (Hrsg.): *Mycobacteria and TB*. Basel : Karger, 2003 (Issues in infectious diseases 2). – ISBN 978-3-8055-7459-4. – OCLC: 834147838 **1.1**
- [98] KAUFMANN, S. H.: How can immunology contribute to the control of tuberculosis? In: *Nature Reviews Immunology* 1 (2001), Nr. 1, 20–30. <http://www.nature.com/nri/journal/v1/n1/abs/nri1001-020a.html> **1.3**
- [99] KAYSER, F. H. ; BÖTTGER, E. C. ; HALLER, O. ; DEPLAZES, P. ; ROERS, A. : Mycobakterien, Tuberkulose. In: *Medizinische Mikrobiologie*. 10. Thieme, 2014. – ISBN 978-3-13-241038-1, S. 273–282 **1.4**
- [100] KEMPKER, R. R. ; KIPIANI, M. ; MIRTSKHULAVA, V. ; TUKVADZE, N. ; MAGEE, M. J. ; BLUMBERG, H. M.: Acquired Drug Resistance in *Mycobacterium tuberculosis* and Poor Outcomes among Patients with Multidrug-Resistant Tuberculosis. In: *Emerging Infectious Diseases* 21 (2015), Jun.,

- Nr. 6, 992–1001. <http://dx.doi.org/10.3201/eid2106.141873>. – DOI 10.3201/eid2106.141873. – ISSN 1080–6040 **1.6**
- [101] KIM, H. J. ; KWAK, H. K. ; LEE, J. ; YUN, Y. J. ; LEE, J. S. ; LEE, M. S. ; MIN, S. Y. ; PARK, S. K. ; KANG, H. S. ; MAENG, Y. H. ; KIM, S. Y. ; KIM, S. Y. ; KOOK, Y. H. ; KIM, Y. R. ; LEE, K. H.: Patterns of pncA mutations in drug-resistant Mycobacterium tuberculosis isolated from patients in South Korea. In: *The International Journal of Tuberculosis and Lung Disease* 16 (2012), Jan., Nr. 1, S. 98–103. <http://dx.doi.org/10.5588/ijtld.10.0739>. – DOI 10.5588/ijtld.10.0739 **6.1**
- [102] KIM, S. J.: Drug-susceptibility testing in tuberculosis: methods and reliability of results. In: *The European Respiratory Journal* 25 (2005), März, Nr. 3, S. 564–569. <http://dx.doi.org/10.1183/09031936.05.00111304>. – DOI 10.1183/09031936.05.00111304. – ISSN 0903–1936 **1.6**, **1.9**, **6.1**, **6.2.2**
- [103] KLEIN, H.-G. ; ROST, I. : *Next Generation Sequencing*. <http://www.medizinische-genetik.de/index.php?id=next-generation-sequencing>. Version: 2016 **1.2**
- [104] KNIEF, C. : Analysis of plant microbe interactions in the era of next generation sequencing technologies. In: *Plant Genetics and Genomics* 5 (2014), 216. <http://dx.doi.org/10.3389/fpls.2014.00216>. – DOI 10.3389/fpls.2014.00216 **1.2**
- [105] KOBOLDT, D. C. ; ZHANG, Q. ; LARSON, D. E. ; SHEN, D. ; MCELLELLAN, M. D. ; LIN, L. ; MILLER, C. A. ; MARDIS, E. R. ; DING, L. ; WILSON, R. K.: VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. In: *Genome Research* 22 (2012), März, Nr. 3, S. 568–576. <http://dx.doi.org/10.1101/gr.129684.111>. – DOI 10.1101/gr.129684.111. – ISSN 1549–5469 **6.2.2**
- [106] KOCH, R. : *Über die Ätiologie der Tuberkulose*. 1982 **1.1**
- [107] KOCH-WESER, D. : Rifampin, New Hope in the Fight against Tuberculosis. In: *New England Journal of Medicine* 283 (1970), Sept., Nr. 12, 655–656. <http://dx.doi.org/10.1056/NEJM197009172831210>. – DOI 10.1056/NEJM197009172831210. – ISSN 0028–4793 **1.1**

- [108] KOHL, T. A. ; DIEHL, R. ; HARMSSEN, D. ; ROTHGÄNGER, J. ; WALTER, K. M. ; MERKER, M. ; WENIGER, T. ; NIEMANN, S. : Whole-Genome-Based Mycobacterium tuberculosis Surveillance: a Standardized, Portable, and Expandable Approach. In: *Journal of Clinical Microbiology* 52 (2014), Jan., Nr. 7, 2479–2486. <http://dx.doi.org/10.1128/JCM.00567-14>. – DOI 10.1128/JCM.00567-14. – ISSN 0095-1137, 1098-660X 1.5, 1.9, 5.1.2
- [109] KÖHLER, W. ; EGGERS, H. J. ; FLEISCHER, B. ; MARRE, R. ; PFISTER, H. ; PULVERER, G. : *Medizinische Mikrobiologie*. 8. Urban & Fischer, 2001. – ISBN 978-3-437-41640-8 1.1
- [110] KONIETZKO, N. (Hrsg.): *Tuberkulose*. Stuttgart : Thieme, 1999. – ISBN 978-3-13-105191-2. – OCLC: 845308834 1.1, 1.6
- [111] KÖSER, C. U. ; ELLINGTON, M. J. ; CARTWRIGHT, E. J. P. ; GILLESPIE, S. H. ; BROWN, N. M. ; FARRINGTON, M. ; HOLDEN, M. T. G. ; DOUGAN, G. ; BENTLEY, S. D. ; PARKHILL, J. ; PEACOCK, S. J.: Routine Use of Microbial Whole Genome Sequencing in Diagnostic and Public Health Microbiology. In: *PLoS Pathogens* 8 (2012), Aug., Nr. 8, e1002824. <http://dx.doi.org/10.1371/journal.ppat.1002824>. – DOI 10.1371/journal.ppat.1002824 6.1
- [112] KÖSER, C. U. ; ELLINGTON, M. J. ; PEACOCK, S. J.: Whole-genome sequencing to control antimicrobial resistance. In: *Trends in Genetics* 30 (2014), Sept., Nr. 9, 401–407. <http://dx.doi.org/10.1016/j.tig.2014.07.003>. – DOI 10.1016/j.tig.2014.07.003. – ISSN 0168-9525 6.1
- [113] KÖSER, C. U. ; FEUERRIEGEL, S. ; SUMMERS, D. K. ; ARCHER, J. A. C. ; NIEMANN, S. : Importance of the Genetic Diversity within the Mycobacterium tuberculosis Complex for the Development of Novel Antibiotics and Diagnostic Tests of Drug Resistance. In: *Antimicrobial Agents and Chemotherapy* 56 (2012), Dez., Nr. 12, 6080–6087. <http://dx.doi.org/10.1128/AAC.01641-12>. – DOI 10.1128/AAC.01641-12. – ISSN 0066-4804 6.1
- [114] KOUL, A. ; ARNOULT, E. ; LOUNIS, N. ; GUILLEMONT, J. ; ANDRIES, K. : The challenge of new drug discovery for tuberculosis. In: *Nature* 469 (2011), Jan., Nr. 7331, 483–490. <http://dx.doi.org/10.1038/nature09657>. – DOI 10.1038/nature09657. – ISSN 0028-0836 1.3

- [115] KULKARNI, S. ; PFEIFER, J. : *Clinical Genomics*. Academic Press, 2014. – ISBN 978-0-12-405173-7. – Google-Books-ID: lwF0AwAAQBAJ **5.2.1**
- [116] KURTZ, S. ; PHILLIPPY, A. ; DELCHER, A. L. ; SMOOT, M. ; SHUMWAY, M. ; ANTONESCU, C. ; SALZBERG, S. L.: Versatile and open software for comparing large genomes. In: *Genome Biology* 5 (2004), R12. <http://dx.doi.org/10.1186/gb-2004-5-2-r12>. – DOI 10.1186/gb-2004-5-2-r12. – ISSN 1474-760X **3.1**
- [117] LEE, K. W. ; LEE, J. M. ; JUNG, K. S.: Characterization of pncA mutations of pyrazinamide-resistant Mycobacterium tuberculosis in Korea. In: *Journal of Korean Medical Science* 16 (2001), Okt., Nr. 5, S. 537–543. – ISSN 1011-8934 **6.1**
- [118] LI, H. ; DURBIN, R. : Fast and accurate short read alignment with Burrows-Wheeler transform. In: *Bioinformatics* 25 (2009), Jul., Nr. 14, 1754–1760. <http://dx.doi.org/10.1093/bioinformatics/btp324>. – DOI 10.1093/bioinformatics/btp324. – ISSN 1367-4803 **1.8, 3.4**
- [119] LI, H. ; HANDSAKER, B. ; WYSOKER, A. ; FENNEL, T. ; RUAN, J. ; HOMER, N. ; MARTH, G. ; ABECASIS, G. ; DURBIN, R. : The Sequence Alignment/Map format and SAMtools. In: *Bioinformatics* 25 (2009), Aug., Nr. 16, 2078–2079. <http://dx.doi.org/10.1093/bioinformatics/btp352>. – DOI 10.1093/bioinformatics/btp352. – ISSN 1367-4803 **3.3, 3.4, 3.5**
- [120] LI, M. ; STONEKING, M. : A new approach for detecting low-level mutations in next-generation sequence data. In: *Genome Biology* 13 (2012), Mai, Nr. 5, S. R34. <http://dx.doi.org/10.1186/gb-2012-13-5-r34>. – DOI 10.1186/gb-2012-13-5-r34. – ISSN 1474-760X **6.2.2**
- [121] LI, R. ; LI, Y. ; FANG, X. ; YANG, H. ; WANG, J. ; KRISTIANSEN, K. ; WANG, J. : SNP detection for massively parallel whole-genome resequencing. In: *Genome Research* 19 (2009), Jun., Nr. 6, S. 1124–1132. <http://dx.doi.org/10.1101/gr.088013.108>. – DOI 10.1101/gr.088013.108. – ISSN 1088-9051 **2.3**
- [122] LI, S. ; LI, R. ; LI, H. ; LU, J. ; LI, Y. ; BOLUND, L. ; SCHIERUP, M. H. ; WANG, J. : SOAPindel: Efficient identification of indels from short paired reads. In: *Genome Research* 23 (2013), Jan., Nr. 1, 195–200. <http://dx.doi.org/10.1101/gr.124488.1>

- doi.org/10.1101/gr.132480.111. – DOI 10.1101/gr.132480.111. – ISSN 1088–9051 **6.1**
- [123] LÖNNROTH, K. ; JARAMILLO, E. ; WILLIAMS, B. G. ; DYE, C. ; RAVI-GLIONE, M. : Drivers of tuberculosis epidemics: The role of risk factors and social determinants. In: *Social Science & Medicine* 68 (2009), Jun., Nr. 12, 2240–2246. <http://dx.doi.org/10.1016/j.socscimed.2009.03.041>. – DOI 10.1016/j.socscimed.2009.03.041. – ISSN 0277–9536 **1.3**
- [124] LUNTER, G. ; GOODSON, M. : Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. In: *Genome Research* 21 (2011), Jan., Nr. 6, 936–939. <http://dx.doi.org/10.1101/gr.111120.110>. – DOI 10.1101/gr.111120.110. – ISSN 1088–9051, 1549–5469 **3.3**
- [125] MALLARD, K. ; MCNERNEY, R. ; CRAMPIN, A. C. ; HOUBEN, R. ; NDLOVU, R. ; MUNTHALI, L. ; WARREN, R. M. ; FRENCH, N. ; GLYNN, J. R.: Molecular Detection of Mixed Infections of Mycobacterium tuberculosis Strains in Sputum Samples from Patients in Karonga District, Malawi. In: *Journal of Clinical Microbiology* 48 (2010), Dez., Nr. 12, 4512–4518. <http://dx.doi.org/10.1128/JCM.01683-10>. – DOI 10.1128/JCM.01683-10. – ISSN 0095–1137 **1.7**
- [126] MANKIEWICZ, E. ; LIIVAK, M. : Phage types of mycobacterium tuberculosis in cultures isolated from Eskimo patients. In: *The American Review of Respiratory Disease* 111 (1975), März, Nr. 3, S. 307–312. <http://dx.doi.org/10.1164/arrd.1975.111.3.307>. – DOI 10.1164/arrd.1975.111.3.307. – ISSN 0003–0805 **1.7**
- [127] MARCO, D. : *Metagenomics: Theory, Methods and Applications*. Horizon Scientific Press, 2010. – ISBN 978–1–904455–54–7. – Google-Books-ID: W_a7plftTUMC **6.2.1**
- [128] MARGARYAN, H. ; RÜSCH-GERDES, S. ; HAYRAPETYAN, A. ; MIRZOYAN, A. : Ethambutol-resistance testing by mutation detection using MTBDRsl. In: *International Journal of Mycobacteriology* (2016), Jul. <http://dx.doi.org/10.1016/j.ijmyco.2016.06.007>. – DOI 10.1016/j.ijmyco.2016.06.007. – ISSN 2212–5531 **6.1**

- [129] MARLOWE, E. M. ; NOVAK-WEEKLEY, S. M. ; CUMPIO, J. ; SHARP, S. E. ; MOMENY, M. A. ; BABST, A. ; CARLSON, J. S. ; KAWAMURA, M. ; PANDORI, M. : Evaluation of the Cepheid Xpert MTB/RIF assay for direct detection of Mycobacterium tuberculosis complex in respiratory specimens. In: *Journal of Clinical Microbiology* 49 (2011), Apr., Nr. 4, S. 1621–1623. <http://dx.doi.org/10.1128/JCM.02214-10>. – DOI 10.1128/JCM.02214-10. – ISSN 1098-660X 1.6, 1.9, 3.3
- [130] MARSHALL, J. : *The SAM Format Specification*. <http://samtools.sourceforge.net/SAM1.pdf>. Version: 2011 2.2.2
- [131] MATSUMOTO, T. ; ANO, H. ; NAGAI, T. ; DANNO, K. ; TAKASHIMA, T. ; TSUYUGUCHI, I. : IS6110 DNA fingerprinting analysis of individually separated colonies of Mycobacterium tuberculosis. In: *Tuberculosis* 85 (2005), Jul., Nr. 4, 207–212. <http://dx.doi.org/10.1016/j.tube.2005.02.001>. – DOI 10.1016/j.tube.2005.02.001. – ISSN 1472-9792 1.5
- [132] MAZARS, E. ; LESJEAN, S. ; BANULS, A. L. ; GILBERT, M. ; VINCENT, V. ; GICQUEL, B. ; TIBAYRENC, M. ; LOCHT, C. ; SUPPLY, P. : High-resolution minisatellite-based typing as a portable approach to global analysis of Mycobacterium tuberculosis molecular epidemiology. In: *Proceedings of the National Academy of Sciences of the United States of America* 98 (2001), Febr., Nr. 4, S. 1901–1906. <http://dx.doi.org/10.1073/pnas.98.4.1901>. – DOI 10.1073/pnas.98.4.1901. – ISSN 0027-8424 1.5
- [133] MCGRATH, M. ; PITTIUS, G. v. ; C, N. ; HELDEN, V. ; D, P. ; WARREN, R. M. ; WARNER, D. F.: Mutation rate and the emergence of drug resistance in Mycobacterium tuberculosis. In: *Journal of Antimicrobial Chemotherapy* 69 (2014), Febr., Nr. 2, 292–302. <http://dx.doi.org/10.1093/jac/dkt364>. – DOI 10.1093/jac/dkt364. – ISSN 0305-7453 1.6
- [134] MCIVOR, A. ; KOORNHOF, H. ; KANA, B. D.: Relapse, re-infection and mixed infections in tuberculosis disease. In: *Pathogens and Disease* 75 (2017), Apr., Nr. 3. <http://dx.doi.org/10.1093/femspd/ftx020>. – DOI 10.1093/femspd/ftx020 1.7, 1.9
- [135] MCKENNA, A. ; HANNA, M. ; BANKS, E. ; SIVACHENKO, A. ; CIBULSKIS, K. ; KERNYTSKY, A. ; GARIMELLA, K. ; ALTSHULER, D. ; GABRIEL, S. ; DALY, M. ; DEPRISTO, M. A.: The Genome Analysis Toolkit: a MapReduce

- framework for analyzing next-generation DNA sequencing data. In: *Genome research* 20 (2010), Sept., Nr. 9, S. 1297–1303. <http://dx.doi.org/10.1101/gr.107524.110>. – DOI 10.1101/gr.107524.110. – ISSN 1549–5469 **2.3**
- [136] MCKEOWN, T. : *The Role of Medicine: Dream, Mirage, or Nemesis?* Princeton University Press, 1979. – ISBN 978–1–4008–5462–2. – Google-Books-ID: xLr_AwAAQBAJ **1.1**
- [137] MEDICINE, N. L.: *BLAST: Basic Local Alignment Search Tool*. <https://blast.ncbi.nlm.nih.gov/Blast.cgi> **3.1**
- [138] MERKER, M. ; BLIN, C. ; MONA, S. ; DUFORET-FREBOURG, N. ; LECHER, S. ; WILLERY, E. ; BLUM, M. G. B. ; RÜSCH-GERDES, S. ; MOKROUSOV, I. ; ALEKSIC, E. ; ALLIX-BÉGUEC, C. ; ANTIERENS, A. ; AUGUSTYNOWICZ-KOPĆ, E. ; BALLIF, M. ; BARLETTA, F. ; BECK, H. P. ; BARRY III, C. E. ; BONNET, M. ; BORRONI, E. ; CAMPOS-HERRERO, I. ; CIRILLO, D. ; COX, H. ; CROWE, S. ; CRUDU, V. ; DIEHL, R. ; DROBNIEWSKI, F. ; FAUVILLE-DUFAUX, M. ; GAGNEUX, S. ; GHEBREMICHAEL, S. ; HANEKOM, M. ; HOFFNER, S. ; JIAO, W.-w. ; KALON, S. ; KOHL, T. A. ; KONTSEVAYA, I. ; LILLEBAEK, T. ; MAEDA, S. ; NIKOLAYEVSKYY, V. ; RASMUSSEN, M. ; RASTOGI, N. ; SAMPER, S. ; SANCHEZ-PADILLA, E. ; SAVIC, B. ; SHAMPUTA, I. C. ; SHEN, A. ; SNG, L.-H. ; STAKENAS, P. ; TOIT, K. ; VARAINE, F. ; VUKOVIC, D. ; WAHL, C. ; WARREN, R. ; SUPPLY, P. ; NIEMANN, S. ; WIRTH, T. : Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. In: *Nature Genetics* 47 (2015), März, Nr. 3, 242–249. <http://dx.doi.org/10.1038/ng.3195>. – DOI 10.1038/ng.3195. – ISSN 1061–4036 **6.2.1**
- [139] MERKER, M. ; KOHL, T. A. ; ROETZER, A. ; TRUEBE, L. ; RICHTER, E. ; RÜSCH-GERDES, S. ; FATTORINI, L. ; OGGIONI, M. R. ; COX, H. ; VARAINE, F. ; NIEMANN, S. : Whole genome sequencing reveals complex evolution patterns of multidrug-resistant Mycobacterium tuberculosis Beijing strains in patients. In: *PloS One* 8 (2013), Nr. 12, S. e82551. <http://dx.doi.org/10.1371/journal.pone.0082551>. – DOI 10.1371/journal.pone.0082551. – ISSN 1932–6203 **F.8**
- [140] MIGLIORI, G. B. ; LODDENKEMPER, R. ; BLASI, F. ; RAVIGLIONE, M. C.:

- 125 years after Robert Koch's discovery of the tubercle bacillus: the new XDR-TB threat. Is "science" enough to tackle the epidemic? In: *European Respiratory Journal* 29 (2007), Nr. 3, 423–427. <http://dx.doi.org/10.1183/09031936.00001307>. – DOI 10.1183/09031936.00001307. – ISSN 0903–1936 **1.2**
- [141] MILLER, F. P. ; VANDOME, A. F. ; JOHN, M. : *Binomial Test*. VDM Publishing, 2010. – ISBN 978–613–2–90006–7. – Google-Books-ID: K8XpYQEACAAJ **5.2.2**
- [142] MORELLI, M. J. ; WRIGHT, C. F. ; KNOWLES, N. J. ; JULEFF, N. ; PATON, D. J. ; KING, D. P. ; HAYDON, D. T.: Evolution of foot-and-mouth disease virus intra-sample sequence diversity during serial transmission in bovine hosts. In: *Veterinary Research* 44 (2013), 12. <http://dx.doi.org/10.1186/1297-9716-44-12>. – DOI 10.1186/1297-9716-44-12. – ISSN 1297–9716 **6.2.2**
- [143] MUSSER, J. M.: Antimicrobial agent resistance in mycobacteria: molecular genetic insights. In: *Clinical Microbiology Reviews* 8 (1995), Okt., Nr. 4, 496–514. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC172873/>. – ISSN 0893–8512 **2.4, 6.2.2**
- [144] NACHEGA, J. B. ; CHAISSON, R. E.: Tuberculosis Drug Resistance: A Global Threat. In: *Clinical Infectious Diseases* 36 (2003), Jan., Nr. Supplement 1, S24–S30. <http://dx.doi.org/10.1086/344657>. – DOI 10.1086/344657. – ISSN 1058–4838, 1537–6591 **1.6**
- [145] NAKATA, N. ; KAI, M. ; MAKINO, M. : Mutation Analysis of Mycobacterial rpoB Genes and Rifampin Resistance Using Recombinant Mycobacterium smegmatis. In: *Antimicrobial Agents and Chemotherapy* 56 (2012), Apr., Nr. 4, 2008–2013. <http://dx.doi.org/10.1128/AAC.05831-11>. – DOI 10.1128/AAC.05831-11. – ISSN 0066–4804 **6.1**
- [146] NAYAK, S. ; ACHARJYA, B. : Mantoux test and its interpretation. In: *Indian Dermatology Online Journal* 3 (2012), Nr. 1, 2–6. <http://dx.doi.org/10.4103/2229-5178.93479>. – DOI 10.4103/2229-5178.93479. – ISSN 2229–5178 **1.1**

- [147] NEELSEN, F. : *Ein casuistischer Beitrag zur Lehre von der Tuberkulose*. 1883 [1.1](#)
- [148] NIELSEN, R. ; PAUL, J. S. ; ALBRECHTSEN, A. ; SONG, Y. S.: Genotype and SNP calling from next-generation sequencing data. In: *Nature reviews. Genetics* 12 (2011), Jun., Nr. 6, 443–451. <http://dx.doi.org/10.1038/nrg2986>. – DOI 10.1038/nrg2986. – ISSN 1471–0056 [2.2.1](#)
- [149] NIEMANN, S. ; RICHTER, E. ; RUSCH-GERDES, S. ; SCHLAAK, M. ; GREINERT, U. : Double infection with a resistant and a multidrug-resistant strain of Mycobacterium tuberculosis. In: *Emerging Infectious Diseases* 6 (2000), Nr. 5, 548–551. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2627962/>. – ISSN 1080–6040 [6.2.1](#)
- [150] NIEMANN, S. ; MERKER, M. ; KOHL, T. ; SUPPLY, P. : Impact of Genetic Diversity on the Biology of Mycobacterium tuberculosis Complex Strains. In: *Microbiology Spectrum* 4 (2016), Nov., Nr. 6. <http://dx.doi.org/10.1128/microbiolspec.TBTB2-0022-2016>. – DOI 10.1128/microbiolspec.TBTB2-0022-2016. – ISSN 2165–0497 [1.2](#), [1.3](#)
- [151] NIEMANN, S. ; RICHTER, E. ; RUSCH-GERDES, S. : Differentiation among Members of the Mycobacterium tuberculosis Complex by Molecular and Biochemical Features: Evidence for Two Pyrazinamide-Susceptible Subtypes of M. bovis. In: *Journal of Clinical Microbiology* 38 (2000), Jan., Nr. 1, 152–157. <http://jcm.asm.org/content/38/1/152>. – ISSN 0095–1137, 1098–660X [6.1](#)
- [152] NIENHAUS, A. (Hrsg.) ; BRANDENBURG, S. (Hrsg.) ; TESCHLER, H. (Hrsg.): *Tuberkulose als Berufskrankheit: ein Leitfaden zur Begutachtung und Vorsorge*. 3., aktualisierte Aufl. Heidelberg : Ecomed Medizin, 2012. – ISBN 978–3–609–16444–1. – OCLC: 774036709 [1.1](#), [1.2](#), [1.2](#)
- [153] OBERDISSE, E. ; HACKENTHAL, E. ; KUSCHINSKY, K. : *Pharmakologie und Toxikologie*. Springer-Verlag, 2013. – ISBN 978–3–642–98030–5. – Google-Books-ID: fySkBgAAQBAJ [1.4](#)
- [154] ORGANIZATION, W. H.: *WHO | The five elements of DOTS*. <http://www.who.int/tb/dots/whatisdots/en/> [1.1](#)

- [155] ORGANIZATION, W. H.: *WHO | WHO End TB Strategy*. http://www.who.int/tb/post2015_strategy/en/ 1.1
- [156] ORGANIZATION, W. H.: *Companion Handbook to the WHO Guidelines for the Programmatic Management of Drug-Resistant Tuberculosis*. World Health Organization (WHO Guidelines Approved by the Guidelines Review Committee). <http://www.ncbi.nlm.nih.gov/books/NBK247420/>. – ISBN 978–92–4–154880–9 1.6
- [157] ORMEROD, L. P.: Multidrug-resistant tuberculosis (MDR-TB): epidemiology, prevention and treatment. In: *British Medical Bulletin* 73-74 (2005), S. 17–24. <http://dx.doi.org/10.1093/bmb/ldh047>. – DOI 10.1093/bmb/ldh047. – ISSN 0007–1420 1.3
- [158] PAPAVENTSIS, D. ; CASALI, N. ; KONTSEVAYA, I. ; DROBNIIEWSKI, F. ; CIRILLO, D. M. ; NIKOLAYEVSKYY, V. : Whole genome sequencing of *Mycobacterium tuberculosis* for detection of drug resistance: a systematic review. In: *Clinical Microbiology and Infection* 23 (2017), Febr., Nr. 2, 61–68. <http://dx.doi.org/10.1016/j.cmi.2016.09.008>. – DOI 10.1016/j.cmi.2016.09.008. – ISSN 1198–743X 1.5, 1.9
- [159] PARTNERSHIP, S. T.: *The global plan to stop TB, 2006-2015*. 2006 1.1, 1.6, 6
- [160] PLINKE, C. ; COX, H. S. ; ZARKUA, N. ; KARIMOVICH, H. A. ; BRACKER, K. ; DIEL, R. ; RÜSCH-GERDES, S. ; FEUERRIEGEL, S. ; NIEMANN, S. : embCAB sequence variation among ethambutol-resistant *Mycobacterium tuberculosis* isolates without embB306 mutation. In: *The Journal of Antimicrobial Chemotherapy* 65 (2010), Jul., Nr. 7, S. 1359–1367. <http://dx.doi.org/10.1093/jac/dkq120>. – DOI 10.1093/jac/dkq120. – ISSN 1460–2091 6.1
- [161] PRIEST, F. ; RAMOS-CORMENZANA, A. ; TINDALL, B. J.: *Bacterial Diversity and Systematics*. Springer Science & Business Media, 2012. – ISBN 978–1–4615–1869–3. – Google-Books-ID: szDjBwAAQBAJ 1.5
- [162] QIAGEN BIOINFORMATICS: *CLC Genomics Workbench*. <https://www.qiagenbioinformatics.com/?qia-storyline=products/clc-genomics-workbench>. Version: 2014 6.1

- [163] RATAN, A. ; OLSON, T. L. ; LOUGHRAN, T. P. ; MILLER, W. : Identification of indels in next-generation sequencing data. In: *BMC Bioinformatics* 16 (2015), 42. <http://dx.doi.org/10.1186/s12859-015-0483-6>. – DOI 10.1186/s12859-015-0483-6. – ISSN 1471-2105 6.1
- [164] REILING, N. ; HOMOLKA, S. ; WALTER, K. ; BRANDENBURG, J. ; NIWINSKI, L. ; ERNST, M. ; HERZMANN, C. ; LANGE, C. ; DIEL, R. ; EHLERS, S. ; NIEMANN, S. : Clade-specific virulence patterns of Mycobacterium tuberculosis complex strains in human primary macrophages and aerogenically infected mice. In: *MBio* 4 (2013), Nr. 4. <http://dx.doi.org/10.1128/mBio.00250-13>. – DOI 10.1128/mBio.00250-13. – ISSN 2150-7511. – PMID: 23900170 1.5, 1.7
- [165] RICHARDSON, E. T. ; LIN, S.-Y. G. ; PINSKY, B. A. ; DESMOND, E. ; BANAEI, N. : First documentation of isoniazid reversion in Mycobacterium tuberculosis. In: *The International Journal of Tuberculosis and Lung Disease: The Official Journal of the International Union Against Tuberculosis and Lung Disease* 13 (2009), Nov., Nr. 11, S. 1347–1354. – ISSN 1815-7920 6.2.2
- [166] RICHTER, M. : *Kriterien der Benutzerfreundlichkeit*. Nov. 1997 4.1, 6.1
- [167] RIDOM BIOINFORMATICS: *Ridom SeqSphere+ - Overview*. <http://www.ridom.de/seqsphere/>. Version: 2013 6.1
- [168] RIE, A. van ; VICTOR, T. C. ; RICHARDSON, M. ; JOHNSON, R. ; SPUY, G. D. d. ; MURRAY, E. J. ; BEYERS, N. ; PITTIUS, N. C. G. ; HELDEN, P. D. ; WARREN, R. M.: Reinfection and Mixed Infection Cause Changing Mycobacterium tuberculosis Drug-Resistance Patterns. In: *American Journal of Respiratory and Critical Care Medicine* 172 (2005), Sept., Nr. 5, 636–642. <http://dx.doi.org/10.1164/rccm.200503-4490C>. – DOI 10.1164/rccm.200503-4490C. – ISSN 1073-449X 1.7, 6.2.1
- [169] RINDER, H. ; MIESKES, K. ; LÖSCHER, T. : Heteroresistance in Mycobacterium tuberculosis. In: *The International Journal of Tuberculosis and Lung Disease* 5 (2001), Apr., Nr. 4, S. 339–345 1.7, 1.7, 6.2.1
- [170] RITTER, C. ; LUCKE, K. ; SIRGEL, F. A. ; WARREN, R. W. ; HELDEN, P. D. ; BÖTTGER, E. C. ; BLOEMBERG, G. V.: Evaluation of the AID

- TB resistance line probe assay for rapid detection of genetic alterations associated with drug resistance in Mycobacterium tuberculosis strains. In: *Journal of Clinical Microbiology* 52 (2014), März, Nr. 3, S. 940–946. <http://dx.doi.org/10.1128/JCM.02597-13>. – DOI 10.1128/JCM.02597-13. – ISSN 1098-660X **3.3**
- [171] ROBERT KOCH-INSTITUT: Bericht zur Epidemiologie der Tuberkulose in Deutschland für 2014. (2015). <http://dx.doi.org/10.17886/rkipubl-2015-005>. – DOI 10.17886/rkipubl-2015-005 **1.2, 1.6**
- [172] RODGER, R ; SHANAGHANY, O.: *Data File Metaformats*. <http://www.catb.org/~esr/writings/taoup/html/ch05s02.html>. Version: 2003 **2.2.4**
- [173] ROTHBERG, J. M. ; HINZ, W. ; REARICK, T. M. ; SCHULTZ, J. ; MILESKI, W. ; DAVEY, M. ; LEAMON, J. H. ; JOHNSON, K. ; MILGREW, M. J. ; EDWARDS, M. ; HOON, J. ; SIMONS, J. F. ; MARRAN, D. ; MYERS, J. W. ; DAVIDSON, J. F. ; BRANTING, A. ; NOBILE, J. R. ; PUC, B. P. ; LIGHT, D. ; CLARK, T. A. ; HUBER, M. ; BRANCIFORTE, J. T. ; STONER, I. B. ; CAWLEY, S. E. ; LYONS, M. ; FU, Y. ; HOMER, N. ; SEDOVA, M. ; MIAO, X. ; REED, B. ; SABINA, J. ; FEIERSTEIN, E. ; SCHORN, M. ; ALANJARY, M. ; DIMALANTA, E. ; DRESSMAN, D. ; KASINSKAS, R. ; SOKOLSKY, T. ; FIDANZA, J. A. ; NAMSARAEV, E. ; MCKERNAN, K. J. ; WILLIAMS, A. ; ROTH, G. T. ; BUSTILLO, J. : An integrated semiconductor device enabling non-optical genome sequencing. In: *Nature* 475 (2011), Jul., Nr. 7356, S. 348–352. <http://dx.doi.org/10.1038/nature10242>. – DOI 10.1038/nature10242. – ISSN 1476-4687 **1.8**
- [174] SANCHEZ-PADILLA, E. ; DLAMINI, T. ; ASCORRA, A. ; RÜSCH-GERDES, S. ; TEFERA, Z. D. ; CALAIN, P. ; TOUR, R. de l. ; JOCHIMS, F. ; RICHTER, E. ; BONNET, M. : High prevalence of multidrug-resistant tuberculosis, Swaziland, 2009-2010. In: *Emerging Infectious Diseases* 18 (2012), Jan., Nr. 1, S. 29–37. <http://dx.doi.org/10.3201/eid1801.110850>. – DOI 10.3201/eid1801.110850. – ISSN 1080-6059 **5.1.4, B.4**
- [175] SANDGREN, A. ; STRONG, M. ; MUTHUKRISHNAN, P. ; WEINER, B. K. ; CHURCH, G. M. ; MURRAY, M. B.: Tuberculosis Drug Resistance Mutation Database. In: *PLoS Med* 6 (2009), Febr., Nr. 2, e1000002. <http://dx.doi.org/10.1371/journal.pmed.0060002>

- [//dx.doi.org/10.1371/journal.pmed.1000002](https://dx.doi.org/10.1371/journal.pmed.1000002). – DOI 10.1371/journal.pmed.1000002 3.2, 6.1
- [176] SASINDRAN, S. J. ; TORRELLES, J. B.: Mycobacterium Tuberculosis Infection and Inflammation: what is Beneficial for the Host and for the Bacterium? In: *Frontiers in Microbiology* 2 (2011), Jan. <http://dx.doi.org/10.3389/fmicb.2011.00002>. – DOI 10.3389/fmicb.2011.00002. – ISSN 1664–302X 1.3
- [177] SCHATZ, A. ; BUGIE, E. ; WAKSMAN, S. A.: Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. 1944. In: *Clinical Orthopaedics and Related Research* (2005), Aug., Nr. 437, S. 3–6. – ISSN 0009–921X 1.1
- [178] SCHLEUSENER, V. : *Analyse von Next-Generation-Sequencing Daten zur Resistenzdetektion bei Tuberkulosebakterien*. März 2014 3.4
- [179] SCHLEUSENER, V. ; KÖSER, C. U. ; BECKERT, P. ; NIEMANN, S. ; FEUERRIEGEL, S. : Mycobacterium tuberculosis resistance prediction and lineage classification from genome sequencing: comparison of automated analysis tools. In: *Scientific Reports* 7 (2017), Apr., 46327. <http://dx.doi.org/10.1038/srep46327>. – DOI 10.1038/srep46327. – ISSN 2045–2322 4, 4.1, 4.3, 4.4, 4.5
- [180] SCHMITT, M. W. ; KENNEDY, S. R. ; SALK, J. J. ; FOX, E. J. ; HIATT, J. B. ; LOEB, L. A.: Detection of ultra-rare mutations by next-generation sequencing. In: *Proceedings of the National Academy of Sciences of the United States of America* 109 (2012), Sept., Nr. 36, S. 14508–14513. <http://dx.doi.org/10.1073/pnas.1208715109>. – DOI 10.1073/pnas.1208715109. – ISSN 1091–6490 6.2.2
- [181] SEKIGUCHI, J.-i. ; MIYOSHI-AKIYAMA, T. ; AUGUSTYNOWICZ-KOPEĆ, E. ; ZWOLSKA, Z. ; KIRIKAE, F. ; TOYOTA, E. ; KOBAYASHI, I. ; MORITA, K. ; KUDO, K. ; KATO, S. ; KURATSUJI, T. ; MORI, T. ; KIRIKAE, T. : Detection of Multidrug Resistance in Mycobacterium tuberculosis. In: *Journal of Clinical Microbiology* 45 (2007), Jan., Nr. 1, 179–192. <http://dx.doi.org/10.1128/JCM.00750-06>. – DOI 10.1128/JCM.00750-06. – ISSN 0095–1137 3.1, 4.3

- [182] SHALER, C. R. ; HORVATH, C. ; JEYANATHAN, M. ; XING, Z. : Within the Enemy's Camp: Contribution of the granuloma to the dissemination, persistence and transmission of *Mycobacterium tuberculosis*. In: *Frontiers in Immunology* 4 (2013), Nr. 30. <http://dx.doi.org/10.3389/fimmu.2013.00030>. – DOI 10.3389/fimmu.2013.00030. – ISSN 1664–3224 **1.3**
- [183] SHAMPUTA, I. C. ; JUGHELI, L. ; SADRADZE, N. ; WILLERY, E. ; PORTAELS, F. ; SUPPLY, P. ; RIGOUTS, L. : Mixed infection and clonal representativeness of a single sputum sample in tuberculosis patients from a penitentiary hospital in Georgia. In: *Respiratory Research* 7 (2006), Jul., Nr. 1, 99. <http://dx.doi.org/10.1186/1465-9921-7-99>. – DOI 10.1186/1465-9921-7-99. – ISSN 1465–9921 **1.7**
- [184] SHENDURE, J. ; JI, H. : Next-generation DNA sequencing. 26 (2008), Okt., Nr. 10, 1135–1145. <http://dx.doi.org/10.1038/nbt1486>. – DOI 10.1038/nbt1486. – ISSN 1087–0156 **1.8**
- [185] SHIN, S. S. ; MODONGO, C. ; NCUBE, R. ; SEPAKO, E. ; KLAUSNER, J. D. ; ZETOLA, N. M.: Advanced Immune Suppression is Associated With Increased Prevalence of Mixed-Strain *Mycobacterium tuberculosis* Infections Among Persons at High Risk for Drug-Resistant Tuberculosis in Botswana. In: *Journal of Infectious Diseases* 211 (2015), Jan., Nr. 3, 347–351. <http://dx.doi.org/10.1093/infdis/jiu421>. – DOI 10.1093/infdis/jiu421. – ISSN 0022–1899, 1537–6613 **6.2.1**
- [186] SILVA, P. E. A. D. ; PALOMINO, J. C.: Molecular basis and mechanisms of drug resistance in *Mycobacterium tuberculosis*: classical and new drugs. In: *J. Antimicrob. Chemother.* 66 (2011), Jan., Nr. 7, 1417–1430. <http://dx.doi.org/10.1093/jac/dkr173>. – DOI 10.1093/jac/dkr173. – ISSN 0305–7453, 1460–2091 **1.2**
- [187] SMITH, I. : *Mycobacterium tuberculosis* Pathogenesis and Molecular Determinants of Virulence. In: *Clin. Microbiol. Rev.* 16 (2003), Jan., Nr. 3, 463–496. <http://dx.doi.org/10.1128/CMR.16.3.463-496.2003>. – DOI 10.1128/CMR.16.3.463–496.2003. – ISSN 0893–8512, 1098–6618. – PMID: 12857778 **1.4, 1.6**
- [188] SPENCER, D. H. ; TYAGI, M. ; VALLANIA, F. ; BREDEMAYER, A. J. ; PFEIFER, J. D. ; MITRA, R. D. ; DUNCAVAGE, E. J.: Performance of Common

- Analysis Methods for Detecting Low-Frequency Single Nucleotide Variants in Targeted Next-Generation Sequence Data. In: *The Journal of Molecular Diagnostics : JMD* 16 (2014), Jan., Nr. 1, 75–88. <http://dx.doi.org/10.1016/j.jmoldx.2013.09.003>. – DOI 10.1016/j.jmoldx.2013.09.003. – ISSN 1525–1578 **6.2.2**
- [189] STARKS, A. M. ; AVILÉS, E. ; CIRILLO, D. M. ; DENKINGER, C. M. ; DOLINGER, D. L. ; EMERSON, C. ; GALLARDA, J. ; HANNA, D. ; KIM, P. S. ; LIWSKI, R. ; MIOTTO, P. ; SCHITO, M. ; ZIGNOL, M. : Collaborative Effort for a Centralized Worldwide Tuberculosis Relational Sequencing Data Platform. In: *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 61Suppl 3 (2015), Okt., S. S141–146. <http://dx.doi.org/10.1093/cid/civ610>. – DOI 10.1093/cid/civ610. – ISSN 1537–6591 **6.1**
- [190] STARZACHER, A. K. ; OTTO-KNAPP, R. ; BÖS, L. ; BAUER, T. : Tuberkulose: Diagnostik und Therapie. In: *Deutsches Zentralkomitee zur Bekämpfung der Tuberkulose* (2014) **1.6**
- [191] STEHLE, T. ; SREERAMULU, S. ; LÖHR, F. ; RICHTER, C. ; SAXENA, K. ; JONKER, H. R. A. ; SCHWALBE, H. : The Apo-structure of the Low Molecular Weight Protein-tyrosine Phosphatase A (MptpA) from *Mycobacterium tuberculosis* Allows for Better Target-specific Drug Development. In: *Journal of Biological Chemistry* 287 (2012), Okt., Nr. 41, 34569–34582. <http://dx.doi.org/10.1074/jbc.M112.399261>. – DOI 10.1074/jbc.M112.399261. – ISSN 0021–9258, 1083–351X **1.4**
- [192] STEINER, A. ; STUCKI, D. ; COSCOLLA, M. ; BORRELL, S. ; GAGNEUX, S. : KvarQ: targeted and direct variant calling from fastq reads of bacterial genomes. In: *BMC Genomics* 15 (2014), Okt., Nr. 1, 881. <http://dx.doi.org/10.1186/1471-2164-15-881>. – DOI 10.1186/1471-2164-15-881. – ISSN 1471–2164 **1.8, 3.2, 4.1, 6.1**
- [193] STEINGART, K. R. ; SCHILLER, I. ; HORNE, D. J. ; PAI, M. ; BOEHME, C. C. ; DENDUKURI, N. : Xpert[®] MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. Version: Jan. 2014. <http://onlinelibrary.wiley.com/doi/10.1002/>

- [14651858.CD009593.pub3/abstract](https://doi.org/10.1002/14651858.CD009593.pub3). In: *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd, 1–46. – DOI: 10.1002/14651858.CD009593.pub3 **1.6**, **1.9**
- [194] STUCKI, D. ; BRITES, D. ; JELJELI, L. ; COSCOLLA, M. ; LIU, Q. ; TRAU-
NER, A. ; FENNER, L. ; RUTAIHWA, L. ; BORRELL, S. ; LUO, T. ; GAO, Q. ;
KATO-MAEDA, M. ; BALLIF, M. ; EGGER, M. ; MACEDO, R. ; MARDASSI,
H. ; MORENO, M. ; VILANOVA, G. T. ; FYFE, J. ; GLOBAN, M. ; THOMAS,
J. ; JAMIESON, F. ; GUTHRIE, J. L. ; ASANTE-POKU, A. ; YEBOAH-MANU,
D. ; WAMPANDE, E. ; SSENGOOBA, W. ; JOLOBA, M. ; BOOM, W. H. ; BA-
SU, I. ; BOWER, J. ; SARAIVA, M. ; VASCONCELLOS, S. E. G. ; SUFFYS,
P. ; KOCH, A. ; WILKINSON, R. ; GAIL-BEKKER, L. ; MALLA, B. ; LEY,
S. D. ; BECK, H.-P. ; JONG, B. C. ; TOIT, K. ; SANCHEZ-PADILLA, E. ;
BONNET, M. ; GIL-BRUSOLA, A. ; FRANK, M. ; PENLAP BENG, V. N.
; EISENACH, K. ; ALANI, I. ; NDUNG’U, P. W. ; REVATHI, G. ; GEHRE,
F. ; AKTER, S. ; NTOUMI, F. ; STEWART-ISHERWOOD, L. ; NTINGINYA,
N. E. ; RACHOW, A. ; HOELSCHER, M. ; CIRILLO, D. M. ; SKENDERS, G.
; HOFFNER, S. ; BAKONYTE, D. ; STAKENAS, P. ; DIEL, R. ; CRUDU, V. ;
MOLDOVAN, O. ; AL-HAJJOJ, S. ; OTERO, L. ; BARLETTA, F. ; CARTER,
E. J. ; DIERO, L. ; SUPPLY, P. ; COMAS, I. n. ; NIEMANN, S. ; GAGNEUX,
S. : Mycobacterium tuberculosis lineage 4 comprises globally distributed
and geographically restricted sublineages. In: *Nature Genetics* 48 (2016),
Dez., Nr. 12, 1535–1543. <http://dx.doi.org/10.1038/ng.3704>. – DOI
10.1038/ng.3704. – ISSN 1061–4036 **6.2.1**
- [195] STUCKI, D. ; GAGNEUX, S. : Single nucleotide polymorphisms in Mycob-
acterium tuberculosis and the need for a curated database. In: *Tuberculosis*
(*Edinburgh, Scotland*) 93 (2013), Jan., Nr. 1, 30–39. <http://dx.doi.org/10.1016/j.tube.2012.11.002>. – DOI 10.1016/j.tube.2012.11.002. – ISSN
1472–9792 **6.2.1**
- [196] STUCKI, D. ; MALLA, B. ; HOSTETTLER, S. ; HUNA, T. ; FELDMANN,
J. ; YEBOAH-MANU, D. ; BORRELL, S. ; FENNER, L. ; COMAS, I. n. ;
COSCOLLÁ, M. ; GAGNEUX, S. : Two New Rapid SNP-Typing Methods
for Classifying Mycobacterium tuberculosis Complex into the Main Phy-
logenetic Lineages. In: *PLoS ONE* 7 (2012), Jul., Nr. 7, e41253. <http://dx.doi.org/10.1371/journal.pone.0171111>. – DOI 10.1371/journal.pone.0171111. – ISSN 1929-0755 **6.2.1**

- [//dx.doi.org/10.1371/journal.pone.0041253](http://dx.doi.org/10.1371/journal.pone.0041253). – DOI 10.1371/journal.pone.0041253 4.2
- [197] SUN, Y.-J. ; LEE, A. S. G. ; WONG, S.-Y. ; HEERSMA, H. ; KREMER, K. ; SOOLINGEN, D. van ; PATON, N. I.: Genotype and phenotype relationships and transmission analysis of drug-resistant tuberculosis in Singapore. In: *The International Journal of Tuberculosis and Lung Disease: The Official Journal of the International Union Against Tuberculosis and Lung Disease* 11 (2007), Apr., Nr. 4, S. 436–442. – ISSN 1027–3719 6.1
- [198] SUN, Y. J. ; LEE, A. S. G. ; WONG, S. Y. ; PATON, N. I.: Analysis of the role of Mycobacterium tuberculosis kasA gene mutations in isoniazid resistance. In: *Clinical Microbiology and Infection* 13 (2007), Aug., Nr. 8, 833–835. <http://dx.doi.org/10.1111/j.1469-0691.2007.01752.x>. – DOI 10.1111/j.1469–0691.2007.01752.x. – ISSN 1198–743X 4.3, 6.1
- [199] SUPPLY, P. ; MAZARS, E. ; LESJEAN, S. ; VINCENT, V. ; GICQUEL, B. ; LOCHT, C. : Variable human minisatellite-like regions in the Mycobacterium tuberculosis genome. In: *Molecular Microbiology* 36 (2000), Mai, Nr. 3, S. 762–771. – ISSN 0950–382X 1.5
- [200] THEISEN, A. ; REICHEL, C. ; RÜSCH-GERDES, S. ; HAAS, W. H. ; ROCKSTROH, J. K. ; SPENGLER, U. ; SAUERBRUCH, T. : Mixed-strain infection with a drug-sensitive and multidrug-resistant strain of Mycobacterium tuberculosis. In: *Lancet (London, England)* 345 (1995), Jun., Nr. 8963, S. 1512. – ISSN 0140–6736 6.2.1
- [201] TOUŠEK, J. ; TRNKA, L. ; HAVEL, A. : Beitrag zum Problem der Resistenzstabilität bei sog. polyresistenten Mycobakterien, erarbeitet von Kranken mit ausgedehnter, chronischer Lungentuberkulose. In: *Beiträge zur Klinik und Erforschung der Tuberkulose und der Lungenkrankheiten* 137 (1968), Febr., Nr. 3, 233–242. <http://dx.doi.org/10.1007/BF02098001>. – DOI 10.1007/BF02098001. – ISSN 0300–9696, 1432–1750 6.2.2
- [202] TSARA, V. ; SERASLI, E. ; CHRISTAKI, P. : Problems in diagnosis and treatment of tuberculosis infection. In: *Hippokratia* 13 (2009), Nr. 1, 20–22. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2633248/>. – ISSN 1108–4189 1.3

- [203] U. KÖSER, C. ; M. BRYANT, J. ; BECQ, J. ; TÖRÖK, M. E. ; ELLINGTON, M. J. ; MARTI-RENOM, M. A. ; CARMICHAEL, A. J. ; PARKHILL, J. ; SMITH, G. P. ; PEACOCK, S. J.: Whole-Genome Sequencing for Rapid Susceptibility Testing of *M. tuberculosis*. In: *The New England journal of medicine* 369 (2013), Jul., Nr. 3. <http://dx.doi.org/10.1056/NEJMc1215305>. – DOI 10.1056/NEJMc1215305. – ISSN 0028–4793 6.2.1
- [204] VALLANIA, F. L. M. ; DRULEY, T. E. ; RAMOS, E. ; WANG, J. ; BORECKI, I. ; PROVINCE, M. ; MITRA, R. D.: High-throughput discovery of rare insertions and deletions in large cohorts. In: *Genome Research* 20 (2010), Dez., Nr. 12, S. 1711–1718. <http://dx.doi.org/10.1101/gr.109157.110>. – DOI 10.1101/gr.109157.110. – ISSN 1549–5469
- [205] VAN RIE, A. ; WARREN, R. ; RICHARDSON, M. ; GIE, R. P. ; ENARSON, D. A. ; BEYERS, N. ; VAN HELDEN, P. D.: Classification of drug-resistant tuberculosis in an epidemic area. In: *The Lancet* 356 (2000), Jul., Nr. 9223, 22–25. [http://dx.doi.org/10.1016/S0140-6736\(00\)02429-6](http://dx.doi.org/10.1016/S0140-6736(00)02429-6). – DOI 10.1016/S0140-6736(00)02429-6. – ISSN 0140–6736 1.6, 1.7
- [206] VILLEMEN, J. A.: *Études sur la tuberculose : preuves rationnelles et expérimentales de sa spécificité et de son inoculabilité*. Paris : J.- B. Baillière et fils <http://archive.org/details/tudessurlatuberc00vill> 1.1
- [207] VYNNYCKY, E. ; FINE, P. E.: The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. In: *Epidemiology and Infection* 119 (1997), Okt., Nr. 2, S. 183–201. – ISSN 0950–2688 6.2.1
- [208] WALKER, T. M. ; KOHL, T. A. ; OMAR, S. V. ; HEDGE, J. ; DEL OJO ELIAS, C. ; BRADLEY, P. ; IQBAL, Z. ; FEUERRIEGEL, S. ; NIEHAUS, K. E. ; WILSON, D. J. ; CLIFTON, D. A. ; KAPATAI, G. ; IP, C. L. C. ; BOWDEN, R. ; DROBNIIEWSKI, F. A. ; ALLIX-BÉGUEC, C. ; GAUDIN, C. ; PARKHILL, J. ; DIEL, R. ; SUPPLY, P. ; CROOK, D. W. ; SMITH, E. G. ; WALKER, A. S. ; ISMAIL, N. ; NIEMANN, S. ; PETO, T. E. A.: Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. In: *The Lancet. Infectious Diseases* 15 (2015), Okt., Nr. 10, 1193–1202. [http://dx.doi.org/10.1016/S1473-3099\(15\)00062-6](http://dx.doi.org/10.1016/S1473-3099(15)00062-6). – DOI 10.1016/S1473-3099(15)00062-6. – ISSN 1473–3099 4.3

- [209] WARREN, R. M. ; VICTOR, T. C. ; STREICHER, E. M. ; RICHARDSON, M. ; BEYERS, N. ; PITTIUS, N. C. v. ; HELDEN, P. D.: Patients with active tuberculosis often have different strains in the same sputum specimen. In: *American Journal of Respiratory and Critical Care Medicine* 169 (2004), März, Nr. 5, S. 610–614. <http://dx.doi.org/10.1164/rccm.200305-7140C>. – DOI 10.1164/rccm.200305-7140C. – ISSN 1073-449X **6.2.1**
- [210] WELTGESUNDHEITSORGANISATION (Hrsg.): *Treatment of tuberculosis: guidelines*. 4th. ed. Geneva : World Health Organization, 2010. – ISBN 978-92-4-154783-3. – OCLC: 742201841 **1.6**
- [211] WIRTH, T. ; HILDEBRAND, F. ; ALLIX-BÉGUEC, C. ; WÖLBELING, F. ; KUBICA, T. ; KREMER, K. ; SOOLINGEN, D. v. ; RÜSCH-GERDES, S. ; LOCHT, C. ; BRISSE, S. ; MEYER, A. ; SUPPLY, P. ; NIEMANN, S. : Origin, Spread and Demography of the Mycobacterium tuberculosis Complex. In: *PLOS Pathogens* 4 (2008), Sept., Nr. 9, e1000160. <http://dx.doi.org/10.1371/journal.ppat.1000160>. – DOI 10.1371/journal.ppat.1000160. – ISSN 1553-7374 **1.5**
- [212] WORLD HEALTH ORGANIZATION: *Global tuberculosis report 2015*. Geneva : World Health Organization, 2015. – ISBN 978-92-4-156505-9 **1.2, 1.6, 2.2, 5.1.3**
- [213] WORLD HEALTH ORGANIZATION: *Global tuberculosis report 2016* <http://apps.who.int/iris/bitstream/10665/250441/1/9789241565394-eng.pdf>. – ISBN 978-92-4-156539-4. – OCLC: 961271202 **1, 1.2, 1.1, 1.2, 1.2, 5.1.4, 6**
- [214] WRIGHT, C. F. ; MORELLI, M. J. ; THÉBAUD, G. ; KNOWLES, N. J. ; HERZYK, P. ; PATON, D. J. ; HAYDON, D. T. ; KING, D. P.: Beyond the Consensus: Dissecting Within-Host Viral Population Diversity of Foot-and-Mouth Disease Virus by Using Next-Generation Genome Sequencing. In: *Journal of Virology* 85 (2011), Jan., Nr. 5, 2266–2275. <http://dx.doi.org/10.1128/JVI.01396-10>. – DOI 10.1128/JVI.01396-10. – ISSN 0022-538X, 1098-5514 **6.2.2**
- [215] WYRES, K. L. ; CONWAY, T. C. ; GARG, S. ; QUEIROZ, C. ; REUMANN, M. ; HOLT, K. ; RUSU, L. I.: WGS Analysis and Interpretation in Clinical and Public Health Microbiology Laboratories: What Are the Require-

- ments and How Do Existing Tools Compare? In: *Pathogens* 3 (2014), Jun., Nr. 2, 437–458. <http://dx.doi.org/10.3390/pathogens3020437>. – DOI 10.3390/pathogens3020437. – ISSN 2076–0817 1.9, 6, 6.1, 6.2.1, 6.2.2, 6.3
- [216] YE, K. ; SCHULZ, M. H. ; LONG, Q. ; APWEILER, R. ; NING, Z. : Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. In: *Bioinformatics* 25 (2009), Jan., Nr. 21, 2865–2871. <http://dx.doi.org/10.1093/bioinformatics/btp394>. – DOI 10.1093/bioinformatics/btp394. – ISSN 1367–4803, 1460–2059 6.1
- [217] ZAHARIA, M. ; BOLOSKY, W. J. ; CURTIS, K. ; FOX, A. ; PATTERSON, D. ; SHENKER, S. ; STOICA, I. ; KARP, R. M. ; SITTLER, T. : Faster and More Accurate Sequence Alignment with SNAP. In: *arXiv:1111.5572 [cs, q-bio]* (2011), Nov. <http://arxiv.org/abs/1111.5572>. – arXiv: 1111.5572 3.5
- [218] ZERBINO, D. R.: Using the Velvet de novo assembler for short-read sequencing technologies. In: *Current protocols in bioinformatics / editorial board, Andreas D. Baxeavanis ... [et al.]* CHAPTER (2010), Sept., Unit–11.5. <http://dx.doi.org/10.1002/0471250953.bi1105s31>. – DOI 10.1002/0471250953.bi1105s31. – ISSN 1934–3396 3.1
- [219] ZETOLA, N. M. ; SHIN, S. S. ; TUMEDI, K. A. ; MOETI, K. ; NCUBE, R. ; NICOL, M. ; COLLMAN, R. G. ; KLAUSNER, J. D. ; MODONGO, C. : Mixed Mycobacterium tuberculosis Complex Infections and False-Negative Results for Rifampin Resistance by GeneXpert MTB/RIF Are Associated with Poor Clinical Outcomes. In: *Journal of Clinical Microbiology* 52 (2014), Jul., Nr. 7, 2422–2429. <http://dx.doi.org/10.1128/JCM.02489-13>. – DOI 10.1128/JCM.02489-13. – ISSN 0095–1137 6.2.2
- [220] ZHANG, Z. ; WANG, Y. ; PANG, Y. ; LIU, C. : Comparison of Different Drug Susceptibility Test Methods To Detect Rifampin Heteroresistance in Mycobacterium tuberculosis. In: *Antimicrobial Agents and Chemotherapy* 58 (2014), Sept., Nr. 9, 5632–5635. <http://dx.doi.org/10.1128/AAC.02778-14>. – DOI 10.1128/AAC.02778-14. – ISSN 0066–4804 6.2.2
- [221] ZHENG, C. ; LI, S. ; LUO, Z. ; PI, R. ; SUN, H. ; HE, Q. ; TANG, K. ; LUO, M. ; LI, Y. ; COUVIN, D. ; RASTOGI, N. ; SUN, Q. : Mixed Infections and Rifampin Heteroresistance among Mycobacterium tuberculo-

sis Clinical Isolates. In: *Journal of Clinical Microbiology* 53 (2015), Jan., Nr. 7, 2138–2147. <http://dx.doi.org/10.1128/JCM.03507-14>. – DOI 10.1128/JCM.03507-14. – ISSN 0095–1137, 1098–660X 1.7



Viola Schleusener

Curriculum Vitae

Persönliche Angaben

Name Viola Schleusener
Geburtsdatum 09.10.1988
Geburtsort Eutin
Anschrift Rolandsweg 6a
24641 Stukenborn
Telefon +491737314007
E-Mail violaschleusener@gmail.com

Akademische Ausbildung

seit 2017 **Forschungszentrum Borstel**, *Postdoktorand.*
2013–2017 **Forschungszentrum Borstel**, *PhD Student.*
2011–2013 **Masterstudium**, *Universität zu Lübeck*, Mathematik in Medizin und Lebenswissenschaften Fachrichtung Biostatistik, Titel der Masterarbeit: *Analyse von Next-Generation-Sequencing Daten zur Resistenzdetektion bei Tuberkulosebakterien.*
Master of Science (1,6)
2008–2011 **Bachelorstudium**, *Universität zu Lübeck*, Mathematik in Medizin und Lebenswissenschaften, Titel der Bachelorarbeit: *Statistische Auswertung der NBI-Studie bei Harnblasenkarzinomen mit SAS.*
Bachelor of Science (2,2)
1999–2008 **Gymnasium**, *Dahlmannschule*, Bad Segeberg, Abschluss Abitur: 1,9.
1995–1999 **Grundschule**, *Astrid-Lindgren-Schule*, Wahlstedt.

Publikationen

- 02-2018 **Schleusener, V., Merker, M., Kohl, T.A. & Niemann, S.**, *Gesamtgenomsequenzierung der Tuberkulose: Anwendungen für molekulare Epidemiologie und Diagnostik*, PNEUMOLOGE, Februar 2018.
doi:10.1007/s10405-017-0156-4
- 02-2018 **Heyckendorf, J., Andres, S., Köser, C. U., Olaru, I. D., Schön, T., Sturegård, E., Beckert, P., Schleusener, V., Kohl, T. A., Hillemann, D., Moradigaravand, D., Parkhill, J., Peacock, S. J., Niemann, S., Lange, C. & Merker, M.**, *What Is Resistance? Impact of Phenotypic versus Molecular Drug Resistance Testing on Therapy for Multi- and Extensively Drug-Resistant Tuberculosis.*, ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, Februar 2018.
doi:10.1128/AAC.01550-17
- 04-2017 **Schleusener, V., Köser, C. U., Niemann, S. & Feuerriegel, S.**, *Mycobacterium tuberculosis resistance prediction and lineage classification from genome data: comparison of automated analysis tools*, SCIENTIFIC REPORTS, 20. April 2017.
doi:10.1038/srep46327
- 03-2017 **Wagner, N-M., Van Aken, C., Butschkau, A., Bierhansl, L., Kellner, P., Schleusener, V., Seggewiss, J., Vollmar, B., Nöldge-Schomburg, G. & Roesner, J.**, *Procalcitonin Impairs Endothelial Cell Function and Viability*, ANESTHESIA AND ANALGESIA, 1. März 2017.
doi: 10.1213/ANE.0000000000001574
- 01-2017 **Ajileye, A., Alvarez, N., Merker, M., Walker, T., Akter, S., Brown, K., Moradigaravand, D., Schön, T., Andres, S., Schleusener, V., Vally Omar, S., Coll, F., Huang, H., Diel, R., Ismail, N., Parkhill, J., de Jong, B. C., Peto, T., Crook, D., Niemann, S., Robledo, J., Smith, G., Peacock, S. & Köser, C.**, *Some synonymous and non-synonymous gyrA mutations in Mycobacterium tuberculosis lead to systematic false-resistance results to fluoroquinolones with the Hain GenoType MTBDRsl assays*, ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, 30. Januar 2017.
doi: 10.1128/AAC.02169-16
- 09-2015 **Kellner, P., Schleusener, V., Bauerfeind, F. & Soukup, J.**, *Influence of different infracardial positions of central venous catheters in hemodynamic monitoring using the transpulmonary thermodilution method*, JOURNAL OF CLINICAL MONITORING AND COMPUTING, 4. September 2015.
doi: 10.1007/s10877-015-9762-z
- 04-2015 **Feuerriegel, S., Schleusener, V., Beckert, P., Kohl, T. A., Miotto, P., Cirillo, D. M., Cabibbe, A. M., Niemann, S. & Fellenberg, K.**, *PhyResSE: a Web Tool Delineating Mycobacterium tuberculosis Antibiotic Resistance and Lineage from Whole-Genome Sequencing Data*, JOURNAL OF CLINICAL MICROBIOLOGY, 8. April 2015.
doi: 10.1128/JCM.00025-15

Danksagungen

Nach vielen Jahren intensiver Arbeit liegt sie nun vor Ihnen: meine Dissertation. Damit ist es an der Zeit, mich bei denjenigen zu bedanken, die mich in dieser herausfordernden, aber auch ungemein lohnenden Phase meiner akademischen Laufbahn begleitet haben. An erster Stelle gilt mein Dank meinem Doktorvater Herr Prof. Dr. Stefan Niemann für seine wissenschaftliche Unterstützung während der gesamten Bearbeitungsphase meiner Dissertation, sowie der Möglichkeit während der Zeit an wertvollen, internationalen Fachkonferenzen teilnehmen zu können.

Allen Mitarbeitern der Molekularen und Experimentellen Mykobakteriologie bin ich sehr dankbar für die gute und zahlreiche Unterstützung, sowie für die konstruktive und angenehme Zusammenarbeit. Ein besonderer Dank geht an die Mädels im Labor, die mich nicht nur mit sehr guten Daten versorgt haben, sondern immer ein Ohr für mich hatten und mich stets unterstützt haben. Auch die vielen nicht-wissenschaftlichen und motivierenden Gespräche haben meine Arbeit unterstützt. Ein spezieller Dank gilt den Kollegen Sven Malm und Christian Utpatel, die mich mit zielführenden Diskussionen und anhaltender Hilfestellung begleitet und unterstützt haben. Für die Unterstützung im Bereich der statistischen Auswertung meiner Daten möchte ich an dieser Stelle noch Jochen Kruppa danken, der spontan und dennoch mit Hingabe die statistischen Belange meiner Arbeit überprüft hat. Zu guter Letzt möchte ich an dieser Stelle auch meiner Familie Pe, Paps, Biene, meiner Oma und Henning für die unermüdliche Stärkung und Motivation danken, sowie für den Rückhalt und die Geduld.