

## Next Generation Sequencing zur Aufklärung von Infektionswegen

**Jörg Linde, Hanka Brangsch, Mostafa Y. Abdel-Gliil, Hosny El-Adawy, Mandy Elschner, Falk Melzer, Katja Mertens-Scholz, Ulrich Methner, Christian Seyboldt, Herbert Tomaso, Gamal Wareth, Silvia Garcia-Soto, Lisa D. Sprague, Heinrich Neubauer**

Bundesforschungsinstitut für Tiergesundheit, Friedrich-Loeffler-Institut (FLI), Institut für bakterielle Infektionen und Zoonosen (IBIZ), Jena

### Zusammenfassung

Die Aufklärung von Infektionswegen ist ein wichtiger Bestandteil der Arbeit der Nationalen Referenzlabore (NRL) und der OIE-Referenzlabore des Instituts für bakterielle Infektionen und Zoonosen (IBIZ), um Übertragungswege von Krankheitserregern nachzuvollziehen, ein adäquates und zeitgerechtes Management von Ausbrüchen zu garantieren und letztendlich Erregerquellen zu eliminieren. Für die Typisierung von bakteriellen Erregern werden traditionell beispielsweise Bio- und Serovar-Bestimmung oder Pulsed-Feld-Gelelektrophorese eingesetzt. Diese Techniken werden aber mittelfristig durch höherauflösende und letztendlich auch günstigere Typisierungsverfahren, basierend auf Ganzgenomsequenzierung mittels Next Generation Sequencing (NGS) und automatisierten bioinformatischen Pipelines, abgelöst werden. Die Interpretation dieser Daten muss allerdings durch erfahrene und sachkundige Spezialisten für die untersuchte Krankheit erfolgen. NGS-basierte Typisierung-Pipelines für die wichtigsten Tierseuchen- und Tierkrankheitserreger, sowie die wichtigen Bakterien des ‚One Health‘-Bereichs, wie z.B. *Acinetobacter*, *Clostridium* oder *Salmonella* spp., wurden am IBIZ validiert und implementiert und stehen frei zur Verfügung ([https://gitlab.com/FLI\\_Bioinfo/WGSBAC](https://gitlab.com/FLI_Bioinfo/WGSBAC)). Diese Tools können auch zur Detektion wichtiger biologischer Marker, beispielsweise für Virulenz, Resistenz gegen Antibiotika und von mobilen genetischen Elementen genutzt werden.

### Einleitung

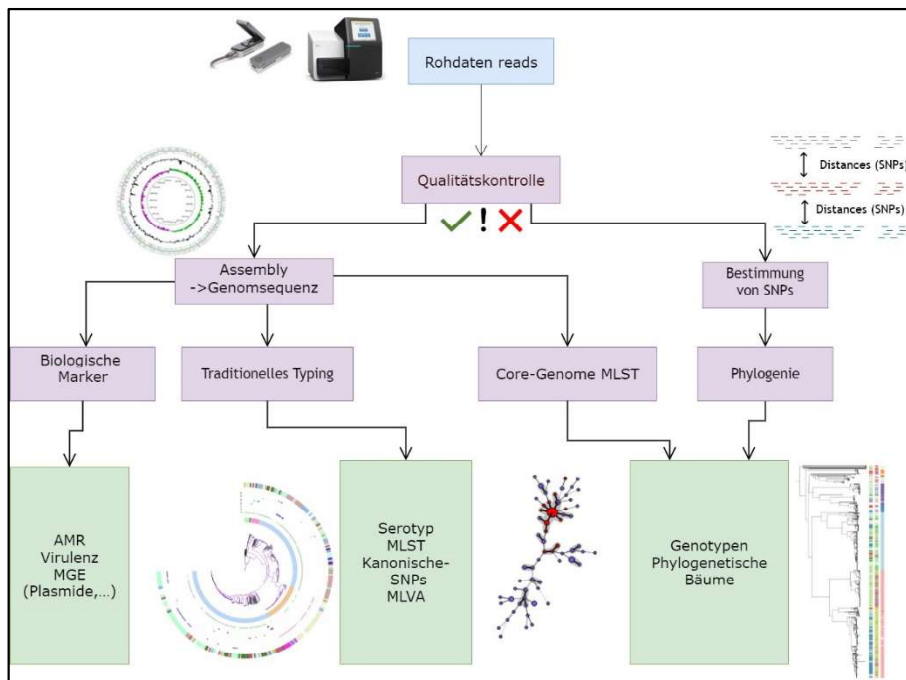
Die Aufklärung von Infektionswegen ist essentiell, um Übertragungswege von Krankheitserregern nachzuvollziehen, Ausbrüche zeitgerecht und adäquat zu managen und um Erregerquellen zu eliminieren. Traditionell für die Typisierung bakterieller Erreger eingesetzte Verfahren sind u.a. die Bio- und Serovar-Bestimmung, Puls-Feld-Gelelektrophorese, PCR mit Sequenzierung der Amplifikate zur Bestimmung kanonischer Einzelnukleotidänderungen (engl.: Single-Nucleotide-Polymorphisms, SNPs) oder zur Multilocus-Sequenztypisierung (MLST) und Multiple-Locus-Variable-Number-Tandem-Repeat Analyse (MLVA). Die aktuelle Notwendigkeit der Entwicklung und Standardisierung von neuen Typisierungsverfahren und Techniken ergibt sich einerseits aus dem rasanten technischen Fortschritt im Bereich der Sequenzieretechnik und Bioinformatik, als auch aus der Umsetzung des EU-Tiergesundheitsrechtsaktes im Bereich des öffentlichen Veterinärwesens. Zusätzlich erfordert die Implementierung der globalen ‚One Health‘-Initiative eine Vergleichbarkeit der Typisierungsergebnisse von Keimen bzw. Resistenzgenen aus Human- und Tiermedizin einerseits und Umweltmedizin andererseits. Durch den hohen Grad der Automatisierung genetischer Analysen, die Entwicklung kostengünstiger Sequenziergeräte, sowie der Verfügbarkeit von geeigneter frei verfügbarer Auswertesoftware und Datenbanken, steht eine wirtschaftliche und akkurate Technik zur Verfügung, die die traditionelle Technik mittelfristig verdrängen wird.

NGS ist ein Überbegriff für verschiedene Methoden zur Sequenzierung von DNS, welche im Vergleich zu Vorgängermethoden (z.B. Sanger-Sequenzierung) schneller und günstiger sind. Während für genetische Untersuchungen beim Menschen (z.B. Vaterschaftstests, pränatale

Diagnostik) oft gezielt einzelne Marker, d.h. kurze Abschnitte der DNS, sequenziert werden, erlauben die vergleichsweise kleinen bakteriellen Genome eine kostengünstige und nahezu vollständige Sequenzierung. Auf dem Markt etabliert hat sich für die Sequenzierung kurzer Sequenzstücke (engl.: reads) beispielsweise die Illumina™-Technologie. Für die Sequenzierung sehr langer Sequenzstücke steht u.a. die Oxford Nanopore™-Technologie (ONT) bereit. Die große Menge an resultierenden Daten (z.B. ca. 30 Gigabyte Rohdaten pro Sequenzierlauf) erfordert eine bioinformatische Auswertung.

Etabliert hat sich die Gesamtgenom-Analyse bakterieller Erreger bereits in der Krankenhaushygiene oder dem Aufklären von lebensmittelbedingten Ausbrüchen. In den USA wird NGS von Erregern bereits routinemäßig eingesetzt.

Die Arbeitsgruppen des IBIZ haben bereits - oder werden zukünftig - validierte Verfahren zur Sequenzierung bakterieller Gesamtgenome und anschließender bioinformatischer Datenauswertung für Tierseuchen- und Tierkrankheitserreger, sowie wichtige Bakterien des ‚One Health‘-Bereichs nationalen und internationalen Institutionen zur Verfügung stellen. Momentan ist es möglich, Rohdaten einer Sequenzierung an die NRLs zu verschicken, um diese am IBIZ analysieren zu lassen und damit die in vielen Institutionen noch nicht etablierten bioinformatischen Abteilungen zu kompensieren.



**Abb. 1:** Workflow; Vereinfachter Ablauf der bioinformatischen Analysen bakterieller Gesamtgenomsequenzierungen.

### Die Bioinformatischen Methoden

Die bioinformatische Auswertung erfolgt in mehreren Schritten, wobei für jeden Schritt verschiedene Softwaretools evaluiert und die jeweils besten Tools miteinander zu einer Pipeline verknüpft werden (Abb.1). Die bioinformatischen Analysen beginnen mit der Qualitätskontrolle der Daten. Dies beinhaltet sowohl die Überprüfung der Qualität und Quantität, als auch die Überprüfung auf Kontaminationen der sequenzierten DNS. Zur Aufklärung von Infektionswegen kommen dann hauptsächlich zwei Verfahren zur Anwendung:

- a) Das Kerngenom MLST (cgMLST) ist die Erweiterung des klassischen MLST (7-9 Gene) auf alle Gene, die im Kerngenom (mobile Elemente ausgeschlossen) einer Bakterienspezies vorkommen. Diese Methode erfordert die *in silico*-Assemblierung (Aneinanderreihung) der sequenzierten DNS-Abschnitte (reads) zum Gesamtgenom. Basierend auf einem vordefinierten Schema wird jedem Gen des assemblierten Gesamtgenoms einer Probe eine Allel-Nummer zugordnet. Die Tabelle von Allel-Nummern für jede Probe wird letztendlich genutzt, um die Proben nach Ähnlichkeit zu gruppieren.
- b) Beim Typisieren, basierend auf SNPs, werden die kurzen reads auf ein Referenzgenom aligniert (gespiegelt), um Einzelnukleotidänderungen (SNPs) zu detektieren. Die Gesamtheit der SNPs aller Proben wird dann mittels statistischer Verfahren (z.B. Maximum-Likelihood) ausgewertet und es werden phylogenetische Bäume erstellt.

Beide Verfahren erlauben eine hochauflösende Typisierung und können Änderungen einzelner Basen (SNPs) bzw. Allele (cgMLST) zwischen zwei Proben detektieren.

Neben den beiden NGS-spezifischen Verfahren, wird die Gesamtgenomsequenz häufig genutzt, um traditionelle Typisierungen *in silico* durchzuführen (Serovar-Bestimmung, MLVA, MLST usw.). Dies dient einerseits zur größeren Typisierung der Proben und andererseits zum Vergleich der sequenzierten Proben mit Probensammlungen, die traditionell untersucht wurden.

Neben der Aufklärung von Infektionswegen, können NGS-Daten auch zur Detektion wichtiger biologischer Marker genutzt werden. Dies sind z.B. Marker für Virulenz, Resistenz gegen Antibiotika oder mobile genetische Elemente (z.B. Plasmide). Die im IBIZ entwickelte Linux-basierte Pipeline zur Analyse bakterieller Gesamtgenomdaten ist frei verfügbar: [https://gitlab.com/FLI\\_Bioinfo/WGSBAC](https://gitlab.com/FLI_Bioinfo/WGSBAC).

Die Interpretation der bioinformatischen Ergebnisse muss allerdings, wie gehabt, durch erfahrene Spezialisten für die untersuchte Krankheit und die jeweiligen Wirtstiere erfolgen.

### Die Erreger

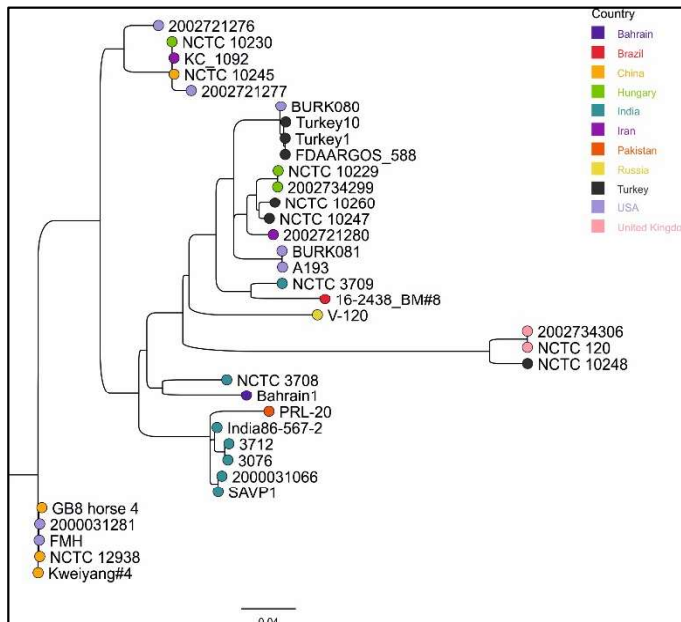
*Acinetobacter (A.) baumannii* ist ein nosokomialer Erreger, der rasch sehr ausgeprägte Antibiotikaresistenzen ausbilden kann. Mittels NGS können diese Resistenzgene detektiert werden (1). Typisierung mittels SNPs und cgMLST ist möglich.

Milzbrand, verursacht durch den ubiquitär vorkommenden Bodenkeim *Bacillus anthracis*, ist eine oft tödlich verlaufende Infektionskrankheit, die vorwiegend pflanzenfressende Tiere betrifft. Hochauflösende Typisierung zur Ausbruchsanalyse erfolgt mittels SNPs und durch das am IBIZ entwickelte cgMLST-Schema (2), welches eine einfache Vergleichbarkeit der Typisierungsergebnisse verschiedener Labore erlaubt. Die traditionelle Typisierung mittels kanonischer SNPs kann *in silico* reproduziert werden. Eine MLVA-basierte Typisierung kann aufgrund der langen Repeat-Sequenzen nur bedingt *in silico* reproduziert werden.

*Brucella melitensis*, *B. abortus* und *B. suis* verursachen die Brucellose, eine nach den Vorgaben des europäischen Tiergesundheitsrechtsakts (EU AHL) bei einigen Nutztierarten zu bekämpfende Tierseuche. *Brucella suis* Biovar 2 ist in Deutschland endemisch und kommt vor allem bei Wildschweinen und Hasen vor. Potentiell besteht auch weiterhin die Gefahr der Einschleppung von *B. melitensis*, *B. suis* Biovar 1 oder *B. abortus*. Bei den Brucellen besitzt die cgSNP-Typisierung ein ausreichend hohes Auflösungsvermögen (3), um die notwendigen epidemiologischen Fragestellungen beantworten zu können. Daneben behalten MLST und MLVA, die ebenfalls *in silico* aus Gesamtgenomsequenzen durchgeführt werden können, für bestimmte Fragestellungen ihre Bedeutung.

*Burkholderia mallei* ist der Erreger des Rotzes der Equiden, kann aber auch auf Fleischfresser oder Kamele und den Menschen übertragen werden. Der Erreger gilt in Europa als ausgerottet, wurde aber bereits mehrmals nach Deutschland eingeschleppt. Eine weltweite Phylogenie, basierend auf NGS-Daten, gibt Rückschlüsse auf länderspezifische Kläden (Abb. 2). Die Typisierung erfolgt traditionell mit MLVA (4), welche auch *in silico* möglich ist. *In silico*-MLST eignet sich nicht für die

Typisierung, da *B. mallei* durch seine Phylogenie innerhalb der Burkholderien nur im uniformen Sequenztyp 40 eingeordnet wird. Hochauflösende Phylogenie erfolgt gegenwärtig mittels SNP-Analysen.



**Abb. 2:** Weltweite Phylogenie von *Burkholderia mallei*, Erreger des Pferderotzes. Phylogenetischer Baum basierend auf einer in silico SNP-Analyse von öffentlich verfügbaren NGS-Daten. Durch Einbeziehung der Metadaten (hier: Herkunftsland des Isolates) können Rückschlüsse gezogen werden, welche Genotypen in bestimmten Regionen dominieren.

*Campylobacter fetus* subsp. *fetus* ist ein Zoonoseerreger. Im Gegensatz dazu verursacht *Campylobacter fetus* subsp. *venerealis* die bovine genitale Campylobakteriose, die zu Aborten bei Rindern und einer hohen wirtschaftlichen Belastung für die Landwirte führt. Die Überwachung und Bekämpfung sind deshalb durch das EU-AHL reguliert. Mittels Sequenzdaten von 283 *Campylobacter fetus*-Stämmen aus 18 Ländern wurde eine stark klonale Gruppe (Klade 1) identifiziert. Alle Stämme aus Klade 1 beherbergen die Insertionssequenz *ISCfe1* und gehören zur Subspezies *venerealis*. Deshalb wurde eine bioinformatische Toolbox für die schnelle Identifizierung dieser Klade auf Basis von NGS-Daten entwickelt, die die Einschränkungen der traditionellen PCR- und MLST-Protokolle für die Subtypisierung von *Campylobacter fetus* kompensiert (5).

Der Rauschbrand, ausgelöst durch *Clostridium chauvoei*, ist eine seuchenhaft und akut verlaufende, infektiöse Gasödemkrankheit, die meist junge Rinder befällt. Infektionen beim Rind mit diesem Erreger sind anzeigepflichtig, bei Schafen und Ziegen meldepflichtig. Eine Feintypisierung mittels cgMLST wurde am IBIZ entwickelt (6) und wird in Zukunft wertvolle Beiträge zu Ausbruchsuntersuchungen liefern.

*Clostridium perfringens* verursacht eine Vielzahl schwerwiegender Infektionen, wobei die Toxinproduktion der zugrundeliegende Mechanismus der Pathogenität bei verschiedenen Wirten ist. Genomische Analysen der Sequenzdaten von über 200 öffentlich zugänglichen *C. perfringens*-Genomen ergaben ein erhebliches Maß an genomischer Variabilität (7).

*Coxiella burnetii*, der Erreger des Q-Fiebers, ist in Deutschland endemisch. Wiederkäuer stellen dabei das Hauptreservoir für Infektionen des Menschen dar. Isolate aus Deutschland können mittels MLVA-Analyse in vier Gruppen (A-D) eingeordnet werden, wobei die meisten Isolate aus Gruppe A oder C sind. Auffällig ist, dass Isolate der Gruppe C fast ausschließlich von Rindern isoliert wurden (8).

Eine NGS-basierte Typisierung ist daher ein wichtiger Faktor bei der Analyse von epidemiologischen Zusammenhängen im Ausbruchgeschehen.

*Francisella tularensis* subsp. *holarctica* wird in Deutschland am häufigsten aus toten Feldhasen isoliert. Durch Kontakt mit infektiösem Material oder (seltener) durch Nahrungsmittel, kann der Erreger die Tularämie (Hasenpest) beim Menschen auslösen. Größere Ausbrüche gab es in den letzten Jahren bei Jägern und Winzern. NGS-basierte Typisierung mittels cgSNPs und cgMLST ist etabliert. Die Detektion von kanonischen SNPs mittels Entscheidungsbäumen spielt weiterhin eine wichtige Rolle. Eine Verbreitungskarte mit hochauflösender Typisierung von über 500 deutschen Isolaten wurde erstellt. Genomisch nahezu identische Isolate finden sich häufig in kleinen geografischen Nischen (z.B. Landkreisen), können in Ausnahmefällen aber auch über weite Distanzen entfernt gefunden werden (9).

Salmonellen verursachen Infektionen und Erkrankungen bei allen landwirtschaftlichen Nutztieren und beim Menschen. NGS-basierte Typisierung mittels cgSNPs und cgMLST ist für Salmonellen etabliert und bietet neben der Analyse phylogenetischer Zusammenhänge bei epidemiologischen Fragestellungen (z.B. Auftreten von neuartigen Erregervarianten) auch die Möglichkeit der Identifikation von Virulenzfaktoren und Resistenzmechanismen gegen Antibiotika und Desinfektionsmittel (10). Darüber hinaus ermöglichen Cluster-Analysen zum geographischen Vorkommen bestimmter *Salmonella*-Serovaren in einer Tierpopulation wichtige Informationen zur Ausbreitung und zur Persistenz über lange Zeiträume und damit Hinweise für Interventionsstrategien (11).

*Yersinia enterocolitica* und *Y. pseudotuberculosis* können bei Menschen und Tieren schwere Infektionen verursachen. *Y. ruckeri*, das die Rotmaulseuche der Salmoniden hervorruft und in Deutschland endemisch vorkommt, spielt aufgrund von z.T. erheblichen wirtschaftlichen Schäden in der Fischzucht eine gewisse Rolle. Für die Typisierung von *Yersinia* spp. wird cgMLST eingesetzt (12).

## Literatur

1. Wareth G, Linde J, Hammer P, Splettstoesser WD, Pletz MW, Neubauer H, Sprague LD. Molecular Characterization of German *Acinetobacter baumannii* isolates and Multilocus Sequence Typing (MLST) analysis based on WGS reveals novel STs. *Pathog.* 2021;10(6):690. doi: 10.3390/pathogens10060690.
2. Abdel-Gliil MY, Chiaverini A, Garofolo G, Fasanella A, Parisi A, Harmsen D, Jolley KA, Elschner MC, Tomaso H, Linde J, Galante D. A whole-genome-based gene-by-gene typing system for standardized high-resolution strain typing of *Bacillus anthracis*. *J Clin Microbiol.* 2021;59(7):e0288920. doi: 10.1128/JCM.02889-20.
3. Khan AU, Melzer F, Sayour AE, Shell WS, Linde J, Abdel-Gliil M, El-Soally SAGE, Elschner MC, Sayour HEM, Ramadan ES, Mohamed SA, Hendam A, Ismail RI, Farahat LF, Roesler U, Neubauer H, El-Adawy H. Whole-Genome Sequencing for tracing the genetic diversity of *Brucella abortus* and *Brucella melitensis* isolated from livestock in Egypt. *Pathog.* 2021;10(6):759.
4. Singha H, Elschner MC, Malik P, Saini S, Tripathi BN, Mertens-Scholz K, Brangsch H, Melzer F, Singh RK, Neubauer H. Molecular typing of *Burkholderia mallei* isolates from equids with glanders, India. *Emerg Infect Dis.* 2021;27(6):1745-1748. doi: 10.3201/eid2706.203232.
5. Abdel-Gliil MY, Hotzel H, Tomaso H, Linde J. Phylogenomic analysis of *Campylobacter fetus* reveals a clonal structure of insertion element *ISCfe1* positive genomes. *Front Microbiol.* 2020;11:585374. doi: 10.3389/fmicb.2020.585374.
6. Thomas P, Abdel-Gliil MA, Eichhorn I, Semmler T, Werckenthin C, Baumbach C, Murmann W, Bodenthin-Drauschke A, Zimmermann P, Schotte U, Galante D, Slavic D, Wagner M, Wieler LH, Neubauer H, Seyboldt C. Genome sequence analysis of *Clostridium chauvoei* strains of European origin and evaluation of typing options for outbreak investigations. *Front Microbiol.* 2021, in press.
7. Abdel-Gliil MY, Thomas P, Linde J, Busch A, Wieler LH, Neubauer H, Seyboldt C. Comparative *in silico* genome analysis of *Clostridium perfringens* unravels stable phylogroups with different genome characteristics and pathogenic potential. *Sci Rep.* 2021;11(1):6756. doi: 10.1038/s41598-021-86148-8.

8. Frangoulidis D, Walter MC, Antwerpen M, Zimmermann P, Janowetz B, Alex M, Böttcher J, Henning K, Hilbert A, Ganter M, Runge M, Münsterkötter M, Splettstoesser WD, Hanczaruk M. Molecular analysis of *Coxiella burnetii* in Germany reveals evolution of unique clonal clusters. *Int J Med Microbiol.* 2014;304(7): 868-876. doi: 10.1016/j.ijmm.2014.06.011.
9. Linde J, Homeier-Bachmann T, Dangel A, Riehm J, Sundell D, Öhrman C, Forsman M, Tomaso, H. Genotyping of *Francisella tularensis* subsp. *holarctica* from hares in Germany. *Microorg.* 2020;5;8(12):1932.
10. García-Soto S, Abdel-Gill M, Tomaso H, Linde J, Methner U. Emergence of multidrug-resistant *Salmonella enterica* subspecies *enterica* serovar infantis of multilocus sequence type 2283 in German broiler farms. *Front Microbiol.* 2020;11:1741. doi: 10.3389/fmicb.2020.01741.
11. García-Soto S, Tomaso H, Linde J, Methner U. *Salmonella enterica* subspecies *enterica* serovar Dublin in German cattle herds - epidemiological analysis using whole-genome sequencing. *Microb. Spectrum.* 2021;in press.
12. Savin C, Criscuolo A, Guglielmini J, Le Guern AS, Carniel E, Pizarro-Cerdá J, Brisse S. Genus-wide *Yersinia* core-genome multilocus sequence typing for species identification and strain characterization. *Microb Genom.* 2019;5(10):e000301. doi: 10.1099/mgen.0.000301.

### **Kontakt**

Prof. Heinrich Neubauer, Bundesforschungsinstitut für Tiergesundheit, Friedrich-Loeffler-Institut, Institut für bakterielle Infektionen und Zoonosen (IBIZ), Jena