

Bundesinstitut für Risikobewertung

# **Anwendung des Whole Genome Sequencing zur Aufklärung von lebensmittelbedingten Krankheitsausbrüchen**

**Application of Whole Genome Sequencing for the Detection of  
Foodborne Disease Outbreaks**

## **Impressum**

BfR Wissenschaft

BfR-Autoren:

Burkhard Malorny, Laura Uelze, Maria Borowiak, Carlus Deneke, Simon H. Tausch, Josephine Grützke, Beatrice Baumann, Katharina Thomas, Karsten Nöckler

Anwendung des Whole Genome Sequencing zur Aufklärung von lebensmittelbedingten Krankheitsausbrüchen

Application of Whole Genome Sequencing for the Detection of Foodborne Disease Outbreaks

Herausgeber:

Bundesinstitut für Risikobewertung

Anstalt des öffentlichen Rechts

Vertreten durch den Präsidenten Professor Dr. Dr. Andreas Hensel

Max-Dohrn-Straße 8–10

10589 Berlin

Telefon: 030 18412 - 0

Fax: 030 18412 - 99099

E-Mail: [pressestelle@bfr.bund.de](mailto:pressestelle@bfr.bund.de)

Aufsichtsbehörde: Bundesministerium für Ernährung und Landwirtschaft

Ust.-IdNr. des BfR: DE 165893448

V.i.S.d.P: Dr. Suzan Fiack

Berlin 2020 (BfR-Wissenschaft 1/2020)

42 Seiten, 2 Abbildungen, 3 Tabellen

**ISBN 978-3-948484-06-4**

**ISSN 1614-3841 (Online)**

*DOI: 10.17590/20200219-132642*

**Inhalt**

<b>Zusammenfassung</b>	<b>5</b>
<b>Abstract</b>	<b>5</b>
<b>1 Einleitung</b>	<b>7</b>
<b>2 Gesamtgenomsequenzierung von Isolaten</b>	<b>9</b>
<b>2.1 WGS Technologien</b>	<b>9</b>
<b>2.2 WGS: Arbeitsabläufe, Geräteausstattung, Qualitätsmanagement</b>	<b>10</b>
<b>2.3 Untersuchung von Sequenzen bakterieller Isolate</b>	<b>13</b>
2.3.1 Sequenzqualitätsüberprüfung und Assemblierung	13
2.3.2 Phylogenetischer Vergleich der Genomsequenzen bei der Aufklärung von lebensmittelbedingen Krankheitsausbrüchen	13
2.3.3 Charakterisierung der Genomsequenzen auf Erregereigenschaften	15
2.3.4 Referenzdatenbanken und bioinformatische Tools für Clusteranalysen	15
<b>3 Interpretation von WGS-Daten im Ausbruchsgeschehen</b>	<b>17</b>
<b>3.1 Genetische Variabilität bakterieller Isolate</b>	<b>17</b>
<b>3.2 Schwellenwerte für die Interpretation von WGS-Daten</b>	<b>18</b>
<b>4 Harmonisierung und Standardisierung</b>	<b>23</b>
<b>5 Austausch von WGS-Daten</b>	<b>25</b>
<b>6 Schlussfolgerung</b>	<b>27</b>
<b>7 Anmerkungen</b>	<b>29</b>
<b>8 Referenzen</b>	<b>31</b>
<b>9 Abbildungsverzeichnis</b>	<b>37</b>
<b>10 Tabellenverzeichnis</b>	<b>39</b>
<b>11 Glossar</b>	<b>41</b>



## Zusammenfassung

Die Gesamtgenomsequenzierung (*engl.* Whole Genome Sequencing [WGS]) ermöglicht es, die genetische Ähnlichkeit von bakteriellen Isolaten mit einem außerordentlich hohen Auflösungsvermögen zu untersuchen. Das Verfahren etabliert sich daher weltweit und sektorenübergreifend (human, veterinär, Lebensmittel und Umwelt) im Rahmen der Untersuchung von Krankheitsausbrüchen und Infektketten. Das WGS eröffnet den zuständigen Behörden und Lebensmittelunternehmern neue Möglichkeiten zur Überwachung von bakteriellen Erregern in Lebensmitteln und zur Verbesserung der Lebensmittelsicherheit und des Verbraucherschutzes. Eine routinemäßige Nutzung des WGS erfordert dabei verschiedene Investitionen zur Schaffung der notwendigen Infrastrukturen einschließlich der Harmonisierung, Standardisierung und Qualitätssicherung für die Generierung, Auswertung und Interpretation der WGS-Daten. Mit diesem Review möchten wir insbesondere Experten von zuständigen Behörden, Untersuchungseinrichtungen der Länder und Lebensmittelunternehmer über den Einsatz der Gesamtgenomsequenzierung im Rahmen der Aufklärung von lebensmittelbedingten Krankheitsausbrüchen informieren. Arbeitsschritte, Auswertungsansätze, die Interpretation von WGS-Daten sowie Herausforderungen in der Standardisierung der Verfahren und dem erforderlichen Austausch von Daten werden dargelegt.

Schlüsselwörter: Whole Genome Sequencing, Next Generation Sequencing, lebensmittelbedingte Krankheitsausbrüche, bakterielle Erreger, Typisierung, Standardisierung, Interpretation

## Abstract

Whole Genome Sequencing (WGS) makes it possible to investigate the genetic similarity of bacterial isolates with extraordinarily high resolution. Therefore, the method is being implemented worldwide and across sectors (human, veterinary, food, and environment) for the investigation of disease outbreaks and infection chains. The WGS opens up new possibilities for the competent authorities and food business operators to monitor bacterial pathogens in food and to improve food safety and consumer protection. Routine use of WGS requires various investments to create the necessary infrastructures, including harmonisation, standardisation and quality assurance for the generation, evaluation and interpretation of WGS data. With this review we inform in particular experts of responsible authorities, laboratories of the federal states and food business operators about the application of whole genome sequencing for the detection of foodborne disease outbreaks. Working steps, approaches for analysis, the interpretation of WGS data, as well as challenges in the standardization of the procedures and the need of data exchange are presented.

Keywords: Whole Genome Sequencing, Next Generation Sequencing, foodborne disease outbreaks, bacterial pathogens, typing, standardization, interpretation



## 1 Einleitung

Lebensmittelbedingte Krankheitsausbrüche verursachen in Deutschland nach wie vor jährlich zahlreiche Erkrankungen. Für das Jahr 2017 sind insgesamt 389 lebensmittelbedingte Krankheitsausbrüche mit mindestens 2277 Erkrankungen, 412 Hospitalisierungen und vier Todesfällen an das Robert Koch-Institut (RKI) und das Bundesamt für Verbraucherschutz und Lebensmittelsicherheit (BVL) übermittelt worden (Rosner et al. 2018). Insbesondere *Campylobacter*-Enteritis-Ausbrüche (38 %) und Salmonellose-Ausbrüche (34 %) sind, mit großem Abstand zu Noroviren-Ausbrüchen (5 %), am häufigsten gemeldet worden. Aber auch Listeriose-Ausbrüche konnten in den vergangenen zwei Jahren durch mit *Listeria monocytogenes* belastete Lebensmittel nachgewiesen werden (Kleta et al. 2017).

Der Anteil der in Zusammenhang mit lebensmittelbedingten Ausbrüchen übermittelten Fälle ist jedoch nur ein Bruchteil der gemäß IfSG übermittelten humanen Erkrankungen. So konnten insgesamt für das Jahr 2017 nur 552 Fälle zu 147 Ausbrüchen durch *Campylobacter* spp. und 817 Fälle zu 133 Ausbrüchen durch *Salmonella* spp. zugeordnet werden (Rosner et al. 2018). Insgesamt sind aber 69.414 *Campylobacter*-Enteritis-Fälle und 14.269 Salmonellosefälle gemeldet worden (RKI 2018). Viele Erkrankungsfälle werden daher als sporadisch angesehen bzw. lassen sich in keinen epidemiologischen Zusammenhang mit dem Verzehr von belasteten Lebensmitteln bringen.

Die Typisierung von Erregerisolaten ist ein wesentlicher Baustein bei der Aufklärung von lebensmittelbedingten Ausbrüchen. Ein Vergleich der genetischen Profile verdächtiger Isolate kann entscheidende Hinweise auf das mit dem Erreger ursächlich belastete Lebensmittel oder sogar auf die ursprüngliche Quelle der Kontamination liefern. Ebenso können erkrankte Patienten eindeutig zu einem Ausbruchsgeschehen zugeordnet werden. Ein sog. Ausbruchsstamm verbreitet sich in der Regel sehr schnell durch den direkten Kontakt von Menschen oder durch die Aufnahme des Erregers über belastete Lebensmittel und in einigen Fällen erfolgt die Verbreitung über größere Gebiete in Form von mehreren Ausbruchsklustern. Der weltweite Handel von Lebensmitteln fördert eine solche Verbreitung. In Verbindung mit verfügbaren Metadaten wie Datum, Ort der Isolierung, Herkunft, Produzent etc. können epidemiologische Untersuchungen im Rahmen von Krankheitsausbrüchen gezielt durchgeführt werden. Steht ein Lebensmittel im Verdacht, einen Krankheitsausbruch verursacht zu haben, können betroffene Patienten genauer nach dem Verzehr des Lebensmittels befragt werden. Ebenso können die Warenkettenströme des Lebensmittels gezielter rückverfolgt werden (Weiser et al. 2016).

Das Auflösungsvermögen der eingesetzten Typisierungsmethoden für den Vergleich von Isolaten auf ihre genetische Verwandtschaft ist dabei für den Erfolg der Untersuchungen entscheidend. In den letzten zwei Jahrzehnten haben insbesondere die Pulsfeldgelelektrophorese (PFGE) und die Multilocus-variable-number-tandem-repeats-Analyse (MLVA) zur Subtypisierung von Isolaten einen bedeutenden Beitrag zur Rückverfolgung und dem Aufspüren von Infektionsquellen geleistet. Ihr Auflösungsvermögen ist jedoch beschränkt, da nur ein Bruchteil der genetischen Information der Isolate durch die Typisierungsmethoden berücksichtigt wird. Phylogenetische Analysen eng verwandter Isolate sind daher weder mit der PFGE noch mit der MLVA möglich.

Mit dem Aufkommen der Gesamtgenomsequenzierung (*engl.* Whole Genome Sequencing [WGS]) in den letzten Jahren ist es möglich geworden, die Basenabfolge eines DNA-Moleküls oder ganzen Genoms eines Erregers zu bestimmen (Heather and Chain 2016; Ronholm et al. 2016). Das höchstmögliche Auflösungsvermögen des WGS, die Verringerung der Kosten des WGS und die Einführung von anwenderfreundlichen Sequenziergeräten in den letzten Jahren macht die Methode für Behörden im Bereich der öffentlichen Gesundheit, Veterinärmedizin und Lebensmittelsicherheit für die Erregertypisierung sehr attraktiv (Kovac et al. 2017; Sekse et al. 2017). Auch spezifische Eigenschaften des Erregers, wie Virulenz-

und Resistenzdeterminanten, können in einem Experiment abgerufen werden. Einige Länder wie Großbritannien, Kanada und die USA haben bereits ein auf WGS basierendes Echtzeit-System für die Überwachung von wichtigen bakteriellen Erregern etabliert, welches für die Untersuchung von Krankheitsausbrüchen und die Überwachung von Kontaminationen in der Lebensmittelproduktion eingesetzt wird (Allard et al. 2016; Ashton et al. 2016; Stevens et al. 2017). Ein solches System ist darauf ausgerichtet, prospektiv auf Grundlage der WGS-Daten Cluster von sehr ähnlichen Isolatsequenzen zu erkennen. Die Kombination dieser Sequenzvergleiche mit den relevanten Metadaten zu den Isolaten und den Warenketteninformationen erlaubt es schließlich, die Übertragungswege zurückzuverfolgen und die Ausbruchsquelle sicher aufzuklären (EFSA BIOHAZ Panel et al. 2019; FAO 2016). Aufgrund der so möglichen frühen Identifizierung eines verdächtigen Lebensmittelvehikels können schneller Maßnahmen ergriffen und somit die Zahl von Erkrankungen in einem Ausbruch verringert werden (WHO 2018). Erfahrungen aus den USA mit der Echtzeit-Überwachung von Listerieninfektionen basierend auf Sequenzdaten von *L. monocytogenes*-Lebensmittel- und -Umweltisolaten aus Monitoringprogrammen zeigen, dass deutlich mehr humane Listeriosecluster und Ausbrüche identifiziert werden konnten. Im Ergebnis der eingeleiteten Maßnahmen konnte pro Ausbruch der Median von sechs auf drei Fälle reduziert werden (Jackson et al. 2016). Eine europäische Studie zu Listerioseausbrüchen kommt zum Schluss, dass europaweite Ausbruchscluster im Durchschnitt mehrere Monate früher erkannt werden, wenn die Daten zentral hinterlegt werden (Van Walle et al. 2018). Auch anderswo konnte gezeigt werden, dass unter Einbeziehung von WGS-Daten die korrekte Einordnung der zu einem Ausbruch gehörenden Isolate als „wahrscheinlich“ oder „nicht wahrscheinlich“ bedeutend gesteigert werden konnte (Jenkins et al. 2019). In einer Studie aus Kanada konnten bereits in einer Kosten-Nutzen-Analyse auch die gesamtökonomischen Vorteile eines solchen Systems für den Erreger *Salmonella* beschrieben werden (Jain et al. 2019).

Inzwischen gibt es internationale Anstrengungen, eine Echtzeit-Überwachung von lebensmittelassoziierten Krankheitserregern zu etablieren und zu standardisieren (Gardy und Loman 2018; Nadon et al. 2017; Taboada et al. 2017). Es gibt jedoch noch einige Herausforderungen zu bewältigen, damit das WGS eine breite Anwendung in der Aufklärung von lebensmittelbedingten Ausbrüchen findet. Nicht nur die Etablierung der Technologie und die damit verbundene Auswertung der Daten ist zu beachten, auch das Datenmanagement spielt für den Erfolg des Verfahrens eine wichtige Rolle. Der Artikel beschreibt die derzeitigen Ansätze der WGS, Arbeitsschritte, Auswertungsmöglichkeiten, die Interpretation von WGS-Daten sowie Herausforderungen in der Standardisierung der Verfahren und dem erforderlichen Austausch von Daten. Dabei wird Bezug auf die Überwachung und die Untersuchung von überregionalen Krankheitsausbrüchen in Deutschland genommen. Erklärungen zu bestimmten Fachausdrücken sind in einem Glossar hinterlegt.



## 2 Gesamtgenomsequenzierung von Isolaten

### 2.1 WGS-Technologien

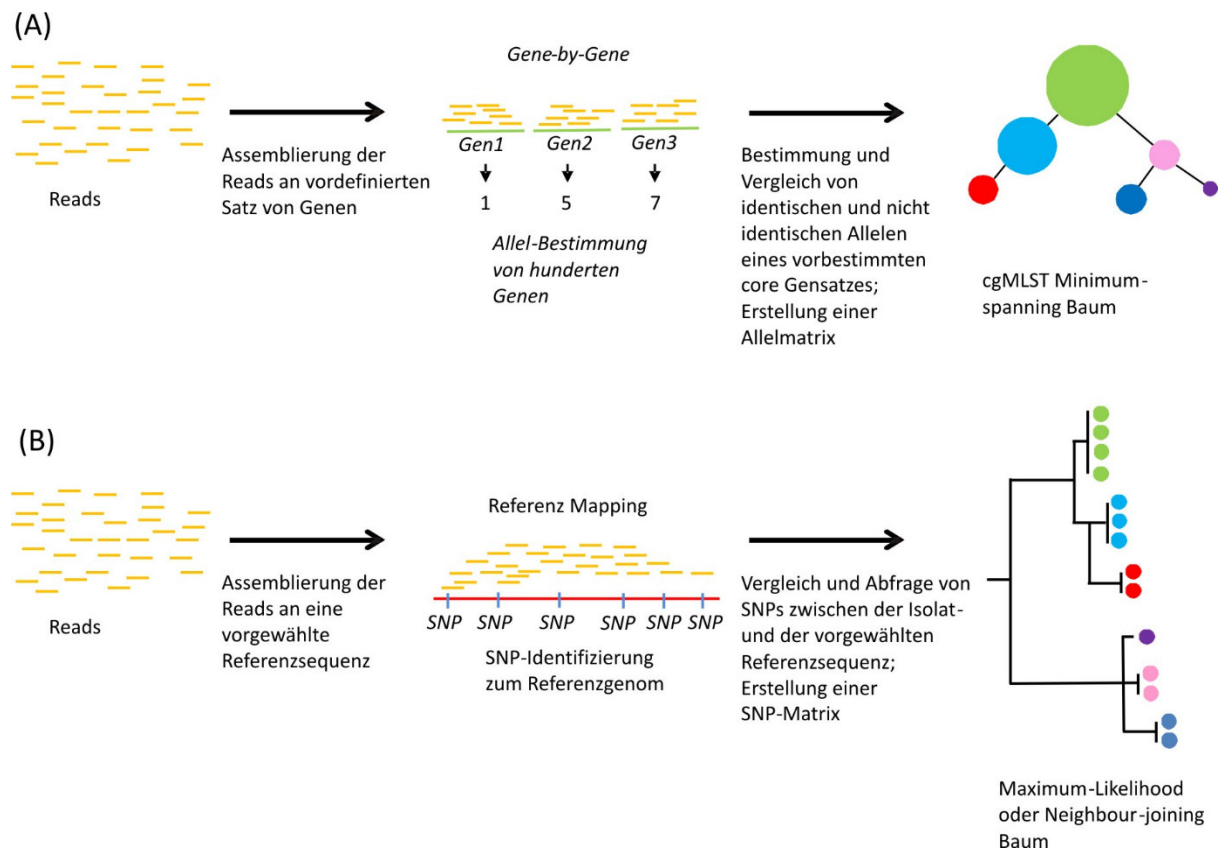
Die WGS basiert auf der „Next Generation Sequencing“-Technologie (NGS, dt. Nächste-Generation-Sequenzierung). NGS-Technologien stellen eine Weiterentwicklung der Sequenzierung der ersten Generation dar, die von Sanger Ende der 1970er-Jahre zur Bestimmung der Basenabfolge von einzelnen kurzen DNA-Molekülen entwickelt wurde (Sanger and Coulson 1975). Der große Fortschritt der Sequenzierung der zweiten Generation basiert auf der massiven millionenfach parallelen Sequenzierung von kurzen (short-read) DNA-Molekülen (50–400 bp Länge) (Cao et al. 2017). Unter Verwendung von bioinformatischen Ansätzen kann aus den erzeugten Sequenzabschnitten die Basenabfolge ganzer Genome bestimmt werden (z.B. *Salmonella* Genom von ca. 5 Megabasen). Nach der kommerziellen Einführung der Sequenzierungsapparaturen der zweiten Generation zu Beginn der 2000er-Jahre folgten Geräte, die in ihrem Umfang auf einem Labortisch platziert werden konnten, in den Anschaffungskosten deutlich gesunken waren und die Vorbereitung der Sequenzierungsproben vereinfachten (Loman et al. 2012a). Dies ermöglichte die breite Anwendung der Technologie in der Mikrobiologie und eröffnete neue Möglichkeiten für Forschung und Routinediagnostik (Didelot et al. 2012; Wilson 2012). Für die Short-read-Genomsequenzierung von Bakterienisolaten hat sich weltweit insbesondere das „sequencing-by-synthesis“-Verfahren, auch Illumina-Technologie genannt, durchgesetzt (<https://www.illumina.com>). Illumina bietet inzwischen eine Reihe von Sequenziergeräten an, die auf den Bedarf unterschiedlicher Labore zugeschnitten sind. Als weitere Short-read-Technologie gibt es das sog. „semi-conductor sequencing“-Verfahren, welches in den Geräten von Thermo Scientific (<https://www.thermofisher.com/de/de/home/life-science/sequencing/next-generation-sequencing.html>) verwendet wird (Cao et al. 2017).

Beim Vergleich bezüglich der Sequenzierungsgenauigkeit weisen beide Technologien derzeit noch geringfügige Unterschiede auf (Loman et al. 2012b; Pallen 2013). Diese technisch bedingten Unterschiede werden sich voraussichtlich durch die künftige Weiterentwicklung von Reagenzien und Softwarelösungen bereinigen lassen. Hinsichtlich der Schnelligkeit der Sequenzierung nehmen sich beide Technologien nicht viel. Im Wesentlichen kommt es bei der Dauer des Laufes auf die Länge der zu sequenzierenden DNA-Moleküle an. Die generierten Sequenzlängen betragen je nach Kitverwendung mit der Illumina Technologie zwischen  $2 \times 75$  bp bis  $2 \times 300$  bp (Paired-end-Sequenzierung) und mit der Ion-Torrent-Technologie zwischen 200 und 400 bp (Single-strand-Sequenzierung).

Einzubeziehen in den Prozess der Sequenzierung sind aber auch die Vorbereitungsschritte der Probe. Eigene Erfahrungen mit der Illumina-Technologie zeigen, dass für die Bearbeitung einer Sequenzierung von kürzeren Sequenzstücken ( $2 \times 75$  bp) von der vorliegenden DNA bis zum ausgewerteten Ergebnis ca. zwei Tage benötigt werden. Für die Erzeugung von längeren Sequenzen, die häufig zu einer besseren Daten-Assemblierung einhergehend mit einer verringerten Contigzahl (Anzahl der bioinformatisch zusammengesetzten Sequenzstücke) führt, werden ca. vier Tage benötigt. Die Anzucht des Erregers für die DNA-Isolierung ist in dieser Betrachtung nicht einbezogen und kann je nach Erreger weitere zwei bis vier Tage erfordern. Für Leser, die an den technologischen Prinzipien der Short-read-Genomsequenzierung interessiert sind, wird auf weitergehende Literatur verwiesen (Besser et al. 2018; Quainoo et al. 2017; Shendure et al. 2017).

Inzwischen sind Sequenziertechnologien der dritten Generation verfügbar, mit denen einzelne DNA-Moleküle sequenziert werden können. Sequenzapparaturen der dritten Generation ermöglichen es, die Basenabfolge sehr langer DNA-Moleküle (ca. 10–50 Kilobasen) am Stück bestimmen zu können (Kwong et al. 2015; Ronholm et al. 2016). Die Long-read-Sequenzierung mit der Oxford-Nanopore-Technologie (ONT) und die „Einzelmolekül-Echtzeit(SMRT)“-Sequenzierung, die von Pacific Biosciences entwickelt wurde, sind in die-

sem Bereich die dominierenden Methoden (Cao et al. 2017; de Lannoy et al. 2017). Die Fehlerrate bei der Erstellung der Basenabfolge durch Sequenziertechnologien der dritten Generation ist vergleichsweise zur zweiten Generation noch hoch, sodass sie bisher nicht für Sequenzvergleiche von Isolaten im Rahmen von Untersuchungen bei Krankheitsausbrüchen aufgrund der Messungengenauigkeit geeignet sind (Quainoo et al. 2017). Sie sind jedoch sehr nützlich, um in Kombination mit Sequenzdaten aus der zweiten Generation geschlossene Genomsequenzen mit hoher Genauigkeit zu generieren. Diese können als Referenzsequenzen für die Ausbruchsanalyse mittels bioinformatischer Verfahren verwendet werden (Abb. 1). Für die meisten relevanten Erreger steht daher auch schon eine Vielzahl geeigneter Referenzsequenzen in öffentlichen Datenbanken zur Verfügung.



**Abb. 1: Methoden der WGS-Datenauswertung im Rahmen von Krankheitsausbrüchen zur Bestimmung der genetischen Ähnlichkeit von Isolatsequenzen. (A) Gene-by-Gene-Annäherung (Allelbestimmung von Genen), (B) Referenzmapping (SNP-Bestimmung anhand eines Referenzgenoms).**

## 2.2 WGS: Arbeitsabläufe, Geräteausstattung, Qualitätsmanagement

Der Arbeitsablauf einer Genomsequenzierung von bakteriellen Isolaten besteht in der Regel aus mehreren Schritten: (i) der DNA-Gewinnung, (ii) der Bibliothekenherstellung unter Verwendung der extrahierten DNA, (iii) der massenhaften Sequenzierung der Bibliothek und (iv) der bioinformatischen Auswertung der Sequenzdaten (Luheshi et al. 2015). Alle Schritte können einen Einfluss auf die Qualität der Ergebnisse haben. Für die Bibliothekenherstellung steht dem Anwender bereits eine größere Palette an Kits von verschiedenen Firmen zur Verfügung (Zähringer 2014). Welche Kitformate für die Anwendungsbereiche des Labors am passendsten sind, sollte ausführlich evaluiert werden (Seth-Smith et al. 2019). Dabei sollte berücksichtigt werden, dass die Sequenzabdeckung über das Genom möglichst gleichmäßig

erfolgt und jede Base des Genoms mit einer ausreichenden Sequenzierungstiefe abgedeckt wird (Sato et al. 2019). Verzerrungen in der Abdeckung können zu falschen oder keinen Ergebnissen führen (Seth-Smith et al. 2019). Ursächlich hierfür können der GC-Gehalt eines Genomabschnittes oder die Generierung von suboptimalen Fragmentgrößen während der Bibliothekenherstellung sein (Aigrain et al. 2016; Tyler et al. 2016). Auch die Methode der DNA-Extraktion kann einen Einfluss auf das Ergebnis haben, insbesondere wenn Isolate Plasmide in ihrem Genom enthalten, deren DNA neben der chromosomalen DNA mitextrahiert werden sollte (Becker et al. 2016).

Mikrobiologische Labore, welche die WGS für die Sequenzierung bakterieller Isolate einführen, benötigen verschiedene Groß- und Kleingeräte. Tabelle 1 gibt ein Beispiel, welche speziellen Geräte notwendig sind. Die Investitionskosten hängen im Wesentlichen von der Anzahl der angestrebten Genomsequenzierungen sowie der bereits bestehenden Labor- und IT-Infrastruktur eines Labors ab. Für die Herstellung der Bibliotheken werden möglicherweise zusätzliche Geräte (z.B. ein Plattenschüttler) benötigt. Bei ca. 500 bis 1000 zu untersuchenden Isolaten im Jahr kann grob eine Investitionssumme von 300.000 € für ein auf Molekular- und Mikrobiologie ausgerichtetes Labor veranschlagt werden. Weitere laufende Kosten entstehen durch die Serviceverträge der Großgeräte, den Unterhalt der IT-Infrastruktur und die entsprechenden Verbrauchsmaterialien für die Sequenzierung (WHO 2018).

**Tab. 1: Benötigte Infrastruktur für ein molekular- bzw. mikrobiologisch arbeitendes Labor zur Genomsequenzierung von bakteriellen Isolaten mittels WGS im Rahmen der Aufklärung von Krankheitsausbrüchen**

Sequenzierungsinfrastruktur	Anmerkung
DNA-Konzentrationsmessung	
Qubit	Fluorometer für die schnelle und genaue Messung der DNA-Konzentration
NanoDrop	Spektralphotometer für die Reinheitsbestimmung der DNA (260/280- und 260/230-Verhältnis) im Mikrovolumenmaßstab
DNA-Bibliothekenerstellung	
Mehrkanalpipetten	Für das effiziente Pipettieren im 96-Well-Maßstab
PCR Cycler	Häufig benötigt, um an fragmentierte DNA Index-Sequenzen für den Sequenzierungsprozess anzuhängen und DNA zu vervielfältigen
Magnetauflagen	Für die Durchführung von Bead-basierten Reinigungsschritten
Plattenzentrifuge	Zum Sammeln von Flüssigkeiten
Optional: Plattenschüttler	Für die Durchmischung von Beads mit Reaktionslösungen
Optional: Fragmentanalyser oder Bioanalyser	Qualitätskontrolle von hergestellten Bibliotheken bezüglich ihrer Konzentration und Fragmentgrößenbestimmung
Optional: Schergerät für DNA	Ultrasonicator für das mechanische Scheren von DNA in kleine Stücke. In Abhängigkeit von dem verwendeten Bibliotheken-Kit ist dieser Schritt notwendig
Sequenzierung	
Sequenzierer	Für die Sequenzierung der DNA-Bibliotheken. Preis und Modell hängen im Wesentlichen von der benötigten Kapazität ab
Optional: unterbrechungsfreie Stromversorgung (USV)	Der Sequenzierer sollte an einer USV angeschlossen sein, um Strompeaks ausgleichen zu können und kurze Stromunterbrechungen zu überbrücken
Datenauswertung	
Datenspeicher/Backup	Speicherung der WGS-Roh- und -Auswertungsdateien. Pro bakterielle Genomsequenz sollten 3 Gigabyte grob kalkuliert werden. Ein Backup-System muss räumlich getrennt von dem eigentlichen Speichersystem vorhanden sein

**Fortsetzung Tab. 1: Benötigte Infrastruktur für ein molekular- bzw. mikrobiologisch arbeitendes Labor zur Genomsequenzierung von bakteriellen Isolaten mittels WGS im Rahmen der Aufklärung von Krankheitsausbrüchen**

Linux Rechner	Für die bioinformatische Datenauswertung. Leistungsstarke Prozessoren sollten die Grundlage eines Linux-Systems sein. Je mehr Rechenleistung, desto schneller die Berechnung (z.B. Assemblierung)
Optional: Windows Rechner mit ausreichender Rechenkapazität	Bei Anwendung von Windows-Programmen werden Windows-Rechner benötigt. Die Rechenkapazität sollte ausreichend sein. Eine Verbindung zu einem Speichersystem muss eingerichtet werden
Kommerzielle und/oder Open-source-Sequenzauswertungsprogramme	Die Anschaffung von bioinformatischen Programmen für die Erstellung von phylogenetischen Bäumen auf SNP-Ebene und/oder cgMLST-Bäumen ist mindestens notwendig. Die Programme müssen auch zur Prüfung von Qualitätsparametern in der Lage sein, sofern kein eigenständiges Programm hierfür benutzt wird
Schnelles Internet	Zum Austausch von WGS-Daten mit Laboren bzw. Behörden muss eine stabile und leistungsfähige Internetanbindung vorhanden sein. Eine 250-Mbit/s-Leitung sollte vorliegen. Die Schnelligkeit sollte sich an dem Bedarf des Labors orientieren

Wie bereits angedeutet, muss die Anschaffung der Infrastruktur für die Genomsequenzierung auch die notwendige Informationstechnik (IT) berücksichtigen (ECDC et al. 2019; Nadon et al. 2017). Leistungsfähige Lösungen für die Datenspeicherung sowie ein Back-up-System müssen angeschafft werden. Bioinformatische Open-source-Programme, die für die Nachbearbeitung der Rohsequenzen und die Auswertung eingesetzt werden, benötigen fast immer Linux-Systeme als Grundlage (Hendriksen et al. 2018; Jagadeesan et al. 2019). Daher ist es notwendig, eine Linux-basierte Infrastruktur zur Verfügung zu haben. Dies stellt häufig für kleinere Labore personell und strukturell eine Herausforderung dar. Zwar gibt es auch Windows-basierte Programme für die Auswertung von Genomsequenzdaten, diese kosten jedoch Lizenzgebühren, enthalten oft überholte Algorithmen und geben keine Möglichkeit zur Weiterentwicklung der Datenauswertung. Der Vorteil ist, dass der Anwender keine tiefgehenden bioinformatischen Kenntnisse besitzen muss. Laboren, die einen höheren Durchsatz und Flexibilität in der Auswertung von Genomsequenzierungen anstreben, wird jedoch empfohlen, Linux-basierte Serversysteme zu etablieren.

Ein WGS-Labor benötigt ein Qualitätsmanagement für die Sequenzierungsabläufe im Labor und die anschließenden bioinformatischen Auswertungen (Hutchins et al. 2019). Da häufig während der Herstellung der Bibliothek noch die Polymerase-Kettenreaktion eingesetzt wird, sind insbesondere die damit in den Laboren verbundenen räumlichen Trennungen zu beachten. Weiterhin muss die Überprüfung der Geräte auf ihre Funktionalität und geforderten Leistungsparameter durch Kontrollen während der Arbeitsschritte sichergestellt sein. Nicht zuletzt ist der richtige Umgang mit dem Probenmaterial zu beachten (Endrullat et al. 2016). Isolate sollten so gelagert und kultiviert werden, dass potenzielle genetische Änderungen durch natürliche Evolutionsprozesse minimiert werden. Dazu gehört das möglichst geringe Passagieren der Kultur, die auf wenige Tage beschränkte Lagerung von Stämmen bei Temperaturen von 4 °C bis 25 °C und das langfristige Archivieren bei tiefen Temperaturen. Der bioinformatischen Analyse kommt in der Qualitätssicherung der erzeugten Genomsequenzen eine besondere Rolle zu (Bogaerts et al. 2019). Mit ihr können Qualitätsparameter des gesamten Sequenzierungslaufs, von allen Sequenzen einer Probe, sowie von jeder individuellen Sequenz geprüft werden. Eine gute Qualität der Sequenzdaten ist schließlich Voraussetzung für verlässliche Vergleichsuntersuchungen.

## 2.3 Untersuchung von Sequenzen bakterieller Isolate

Rohsequenzen, die zu einem Isolat hergestellt wurden, sollten idealerweise direkt nach der Sequenzierung immer in einer Datei verpackt vorliegen. Geräte von Illumina verpacken die Daten bei der Anwendung der üblicherweise benutzten Paired-end-Sequenzierungstechnologie in zwei Dateien. Die Dateien enthalten alle zum Isolat zugeordneten Rohsequenzen. Eine typische Datei hat eine Größe von 0,25 bis 1 Gigabyte und enthält ca. 1 bis 2 Millionen Sequenzstücke. Rohsequenzen, die im Rahmen eines Krankheitsausbruchs von Patienten- und Lebensmittelisolaten generiert wurden, werden zunächst vor einem Vergleich auf ihre genetische Ähnlichkeit einer Qualitätsüberprüfung und Vorbearbeitung unterzogen (Bogaerts et al. 2019).

### 2.3.1 Sequenzqualitätsüberprüfung und Assemblierung

Zur Aufbereitung der Rohsequenzen werden die zum Isolat gehörenden Sequenzen getrimmt und assembliert (Carrico et al. 2018). Die Qualitätsprüfung der Sequenzdaten erfolgt schließlich durch Abfrage wesentlicher Qualitätsparameter, die meist in Form eines automatisch erstellten Assembly Reports aufgelistet werden. Geprüft wird häufig die erwartete Übereinstimmung der vorhergesagten Spezies, die Datenmenge bzw. die Abdeckungstiefe (Coverage) des Assembly, die Anzahl der Sequenzabschnitte (Contigs) sowie die Länge des Gesamt-Assembly, die der erwarteten Länge des Genoms entsprechen sollte (Bogaerts et al. 2019). Die dabei zur Anwendung kommenden Grenzwerte werden auch in einem internationalen ISO-Standard erarbeitet (siehe Kap. 4). Aufgrund der unterschiedlichen Basenzusammensetzung der Erreger-DNA, ihrer Genomgröße und der Variation der Ergebnisse in Abhängigkeit von der angewendeten Sequenzierertechnologie ist es nicht sinnvoll, Qualitätsgrenzwerte generell festzulegen, denn diese sollten sich auch nach den in einem Labor etablierten Sequenzierabläufen und Erfahrungen richten. Tabelle 2 gibt einige Beispiele für Qualitätsparameter mit möglichen Grenzwerten für verschiedene Erreger an.

**Tab. 2: Mindestens anzustrebende Qualitätsparameter der Sequenzierung für ausgesuchte über das Lebensmittel übertragbare Erreger**

	Genomgröße in Mio. Basenpaare	Anzahl sequen- zierende Basen pro Probe (in Mio.)	Abdeckung nach Assem- blierung der Reads	Anzahl der Contigs	Gesamte as- semblierte Län- ge in Mio. Basenpaare
<i>Campylobacter</i>	1,7	> 80	> 40	20–55	1,6–1,8
<i>Listeria</i>	3,0	> 90	> 30	15–45	2,8–3,1
<i>E. coli</i>	5,1	> 200	> 40	150–350	4,9–5,3
<i>Salmonella</i>	4,8	> 150	> 30	50–150	4,6–5,1

### 2.3.2 Phylogenetischer Vergleich der Genomsequenzen bei der Aufklärung von lebensmittelbedingten Krankheitsausbrüchen

Derzeit gibt es eine Reihe verschiedener Ansätze, Genomsequenzen im Rahmen einer Ausbruchsuntersuchung miteinander zu vergleichen (Schürch et al. 2018; Taboada et al. 2017) (Abb. 1). Die gängigsten Ansätze sind: (i) Kerngenom MLST (cgMLST) unter Berücksichtigung von Tausenden von Genen, die in den meisten Isolaten einer Spezies oder Gattung vorhanden sind, (ii) Gesamtgenom MLST (wgMLST) unter Berücksichtigung aller Gene einschließlich der variablen Zusatzgene einer Spezies und (iii) die Referenzkartierung von Einzelnukleotid-Polymorphismen (SNP) (Kovac et al. 2017; Pérez-Losada et al. 2018). Bei der cgMLST/wgMLST wird die DNA-Sequenzvariation in einem Satz von festgelegten Genen bestimmt (Maiden 2006; Maiden et al. 2013). Die Typisierung der Isolate erfolgt anhand ihres

spezifischen Allelprofils, das sich schließlich in ihrem entsprechenden Sequenztyp (ST) widerspiegelt. Während die MLST auf einer „*gene-by-gene*“-basierten Analyse beruht (Gen-für-Gen-Annäherung), beruht die SNP-Analyse auf einem Vergleich der Einzelpunktmutationen der Isolatsequenzen zu einem vorher festgelegten, nahe verwandten Referenzgenom. Diese Art der Analyse nennt man „referenzbasiertes Mapping“ (Carrico et al. 2018). In diesem Verfahren werden keine zu analysierenden Gene festgelegt. So können auch SNPs außerhalb von gencodierenden Sequenzen berücksichtigt werden.

Während bei SNP-basierten Clustering-Verfahren alle Unterschiede, die zum Referenzgenom auftreten, zu erfassen sind, werden bei der cgMLST/wgMLST nur Allelunterschiede ohne Unterscheidung der Anzahl und Art der Mutationen zwischen den zu vergleichenden Isolaten berücksichtigt. Dies kann dazu führen, dass mit der SNP-Analyse ein höheres Auflösungsvermögen als mit der cgMLST/wgMLST-Analyse erreicht wird (Halbedel et al. 2018). Ein Vorteil von cgMLST/wgMLST ist die Anwendung von Nomenklatorschemata, die es ermöglichen, eine eindeutige Identifizierung der zu untersuchenden Sequenz zu erstellen, die leicht kommuniziert werden kann, ohne Rohsequenzen austauschen zu müssen. Voraussetzung dafür ist jedoch die Benutzung einer zentral verwalteten Nomenklatur, die fortlaufend aktualisiert wird (Nadon et al. 2017). CgMLST/wgMLST erfordern prinzipiell etwas weniger Aufwand bei der Rechenleistung im Vergleich zu SNP-basierten Verfahren (Nadon et al. 2017). Durch die verschiedenen bioinformatischen Möglichkeiten, die Qualität der SNPs zu bewerten und daraus schließlich eine Auswahl von SNPs zu treffen, die in einer SNP-Analyse berücksichtigt werden, müssen immer die Rohsequenzen aller zu vergleichenden Isolate vorliegen. SNP-Analysen können nur für relativ eng verwandte Genome sinnvoll angewendet werden, also z.B. innerhalb eines Serovars oder einer MLST-Gruppe. Ein „Vorclustering“ von mit einer SNP-Analyse zu vergleichenden Genomsequenz ist daher sinnvoll. Für beide Verfahren (SNP bzw. cgMLST/wgMLST) gilt grundsätzlich, dass Ergebnisse nur dann vergleichbar sind, wenn sie mit den gleichen oder äquivalenten Softwareprogrammen bzw. identischen Softwarealgorithmen sowie identischen Parametereinstellungen generiert wurden. Bei SNP-Analysen müssen zudem die exakt gleichen Referenzen benutzt werden, da es sonst zu einem unterschiedlichen Ergebnis in der SNP-Identifikation kommen kann (Carrico et al. 2018).

Sowohl cgMLST/wgMLST- als auch SNP-basierte Analysen sind grundsätzlich für die Bestimmung der Unterschiede zwischen Isolaten im Rahmen von Ausbruchsuntersuchungen gut geeignet und führen zu ähnlichen Ergebnissen (Halbedel et al. 2018; Pearce et al. 2018). Beide Analyseverfahren erlauben die Einteilung von Genomen in Subtypen (Clustertypen) und eine Bewertung, ob zwei Genome genetisch (nahezu) identisch sind. Für die Untersuchung der Phylogenie kann die Verwendung von cgMLST- oder SNP-Verfahren robuste Analysen liefern, da sie nur Regionen des Genoms umfassen, die in allen Stämmen vorhanden sind. WgMLST kann aber durch die zusätzliche Berücksichtigung der variablen Zusatzgene einer Spezies zu einer höheren Auflösung führen (Chen et al. 2017). Die Anwendung von cgMLST/wgMLST ist geeigneter, wenn mehrere Benutzer gleichzeitig jedes neue Isolat, das einer gemeinsamen Datenbank hinzugefügt wird, analysieren müssen (z.B. im Ausbruchsfall), insbesondere wenn die Sequenzinformationen nicht öffentlich zugänglich sind (Jagadeesan et al. 2019).

SNP- und cgMLST/wgMLST-Ansätze bewerten genetische Variationen auf etwas unterschiedliche Weise und sollten als komplementär angesehen werden. Insbesondere wenn eine Methode allein keine eindeutige Antwort liefert, sollte der Datensatz zur besseren Beurteilung mit beiden Analyseverfahren ausgewertet werden.

### 2.3.3 Charakterisierung der Genomsequenzen auf Erregereigenschaften

Die o.g. Typisierungsmethoden lassen zunächst keine weitere Aussage zur Zusammensetzung, zum Aufbau und zu den Eigenschaften eines bakteriellen Genoms zu. WGS-Daten lassen sich aber auch für Untersuchungen von Isolaten hinsichtlich ihres Genmaterials und daraus abgeleitete funktionelle Eigenschaften nutzen. So können Virulenzgene, die zur Pathogenität eines Erregers beitragen, Resistenzgene, die für die Unempfindlichkeit gegenüber antimikrobiellen Substanzen und Desinfektionsmitteln verantwortlich sind, oder Gene, die bei der Bildung von Biofilmen eine Rolle spielen, identifiziert werden (Ronholm et al. 2016). Determinanten, die für die Mobilität von Genen eine Rolle spielen, können zur Untersuchung von dynamischen Veränderungen eines Genoms genutzt werden. Diese Informationen ermöglichen es schließlich, bakterielle Stämme hinsichtlich ihres Verhaltens in Mensch, Tier, Umwelt bzw. im Lebensmittel zu bewerten (Luheshi et al. 2015). Die Identifizierung von spezifischen Genen, z.B. in epidemisch aufkommenden Klonen, kann auch für die schnelle Entwicklung von PCR-Schnelltests zur Detektion eines Stammes genutzt werden. In Ausbrüchsfällen können so bei Bedarf verdächtige Proben in einem hohen Durchsatz, kostengünstig untersucht werden, und das auch in Laboratorien, denen WGS nicht zur Verfügung steht (Bielaszewska et al. 2011).

### 2.3.4 Referenzdatenbanken und bioinformatische Tools für Clusteranalysen

Für die Durchführung von Sequenzvergleichen bakterieller Isolate werden bioinformatische Algorithmen und Datenbanken benötigt, die in Auswertungsprogramme integriert werden bzw. darauf zugreifen können. Für die Erstellung von cgMLST/wgMLST-Daten ist ein Schema notwendig, das die zu untersuchenden Gene für die Allelprofilbestimmung festlegt und neuen Allelen eine definierte Nummer zuweist (Maiden et al. 2013). Für eine Auswahl von bedeutenden Erregern (Spezies) liegen bereits öffentlich zugängliche Datenbanken vor, die sowohl WGS-Daten speichern als auch analysieren (Jagadeesan et al. 2019). Diese Datenbanken können meist online über ein Web-Interface-Tool benutzt werden. Prominente Datenbanken mit öffentlich hinterlegten Schemata sind z.B. Enterobase für die Spezies/Genera *Escherichia/Shigella*, *Salmonella enterica*, *Yersinia*, *Vibrio*, *Helicobacter*, *Clostridioides* und *Moraxella* (<https://enterobase.warwick.ac.uk/>), pubMLST insbesondere für *Campylobacter jejuni* und *C. coli* (<https://pubmlst.org/campylobacter/>) sowie PasteurMLST für *L. monocytogenes* (<http://bigsd.pasteur.fr/listeria>). Aufgrund ihrer Bedeutung und Verbreitung finden solche öffentlichen Schemata auch in kommerziellen Programmen Verwendung. Datenbanken können aber auch in Open-source-Programmen hinterlegt werden, sodass Lizenz- oder Analysegebühren an kommerzielle Hersteller nicht anfallen. Für selten vorkommende Spezies können ad hoc cgMLST/wgMLST-Schemata entwickelt werden. Dies ist sinnvoll, wenn öffentliche Referenzschemata nicht zur Verfügung stehen.

Für einen Vergleich von cgMLST-Daten zwischen Laboratorien ist es notwendig, sich auf ein Schema zu der entsprechenden Spezies zu einigen (Nadon et al. 2017). Die Festlegung auf ein Schema pro Spezies ist auch auf internationaler Ebene noch nicht abgeschlossen. Während für einige Spezies (z.B. *Salmonella*) die Anwendung bestimmter Schemata schon relativ gut harmonisiert ist, gibt es für andere Spezies (z.B. *L. monocytogenes*) noch Abstimmungsbedarf. Dies führt dazu, dass Daten zwischen Laboratorien, die unterschiedliche Schemata anwenden, nicht direkt verglichen werden können. Jedoch wurde gezeigt, dass z.B. im Fall *L. monocytogenes* die beiden üblicherweise verwendeten cgMLST-Schemata zu sehr hohen übereinstimmenden Ergebnissen führen (Van Walle et al. 2018).

Bioinformatische Tools und Auswertungspipelines wurden in den letzten Jahren verstärkt entwickelt. Über online Webdienste können viele Auswertungen durchgeführt werden. Dies setzt das Hochladen der zu untersuchenden Sequenz voraus. Das „Center für Genomische Epidemiologie“ der Dänischen Technischen Universität (DTU) in Kopenhagen bietet eine Reihe von Se-

quenzauswertungsmöglichkeiten an, die zur Untersuchung der Sequenzdaten herangezogen werden können, ohne dass diese veröffentlicht werden müssen. Zur Verfügung stehen Tools zur Resistenzgen- und Virulenzgenbestimmung, aber auch Typisierungstools (MLST, cgMLST, Plasmide, Spezies, Serotypie; (<http://www.genomicepidemiology.org/>)). Auch die Plattform PATRIC (<https://www.patricbrc.org/>) ist für den Webanwender hinsichtlich der Analysemöglichkeiten sehr ergiebig. Die bei Online-Webdiensten hinterlegten Softwareprogramme stehen größtenteils der wissenschaftlichen Gemeinschaft als Open-source-Anwendungen zur Verfügung. Bioinformatiker können diese in Pipelines verpacken und automatische Auswertungsprozesse programmieren (Carrico et al. 2018). Eine Online-Anwendung der Tools ist so nicht mehr erforderlich. Das Studienzentrum für Genomsequenzierung und -analyse am Bundesinstitut für Risikobewertung hat eine Pipeline entwickelt, die automatisiert Qualitätsparameter für vom Sequenziergerät erzeugte Rohsequenzen erstellt. Der Anwender kann schließlich anhand der Werte entscheiden, ob die Sequenz qualitativ ausreicht, um einer weiteren Pipeline zur Charakterisierung und Typisierung zugeführt zu werden. Diese semi-automatisierte Vorgehensweise ermöglicht die Analyse von vielen hunderten bis tausenden Genomsequenzen der unterschiedlichen Erreger unabhängig von der Geschwindigkeit der anliegenden Internetverbindung.

Da solche Programme grundsätzlich nur mit dem Betriebssystem Linux kompatibel sind und Befehle hauptsächlich über die Kommandozeile ausgeführt werden, benötigt die Etablierung frei verfügbarer Open-source-Pipelines für die Analyse von Genomsequenzen ein Mindestmaß an bioinformatischer Expertise. Ist die Expertise nicht verfügbar bzw. eine auf Linux ausgerichtete IT-Infrastruktur nicht vorhanden, kann für eine weitergehende Analyse auch auf kommerzielle Windows-basierte Programme zurückgegriffen werden. Nähere zusammenfassende Information zu häufig angewandten Open-source- und kommerziellen Softwaretools für die Genomsequenzcharakterisierung und -typisierung findet man in der Literatur z.B. unter den angegebenen Referenzen (ECDC et al. 2019; Hendriksen et al. 2018; Jagadeesan et al. 2019; Quainoo et al. 2017; WHO 2018).



### 3 Interpretation von WGS-Daten im Ausbruchsgeschehen

Die WGS-Analyse liefert grundsätzlich Informationen darüber, inwieweit Isolate genetisch verwandt sind. Eine engere genetische Verwandtschaft bedeutet aber nicht unbedingt, dass ein klinischer Fall direkt mit einem bestimmten Lebensmittel in Zusammenhang steht. Es ist daher unerlässlich, dass epidemiologische Informationen zum Isolat und dem im Verdacht stehenden Lebensmittel und idealerweise auch Informationen zu den Produktionsbedingungen vorliegen, um die phylogenetischen WGS-Daten richtig zu interpretieren (EFSA BIOHAZ Panel et al. 2019; Jagadeesan et al. 2019).

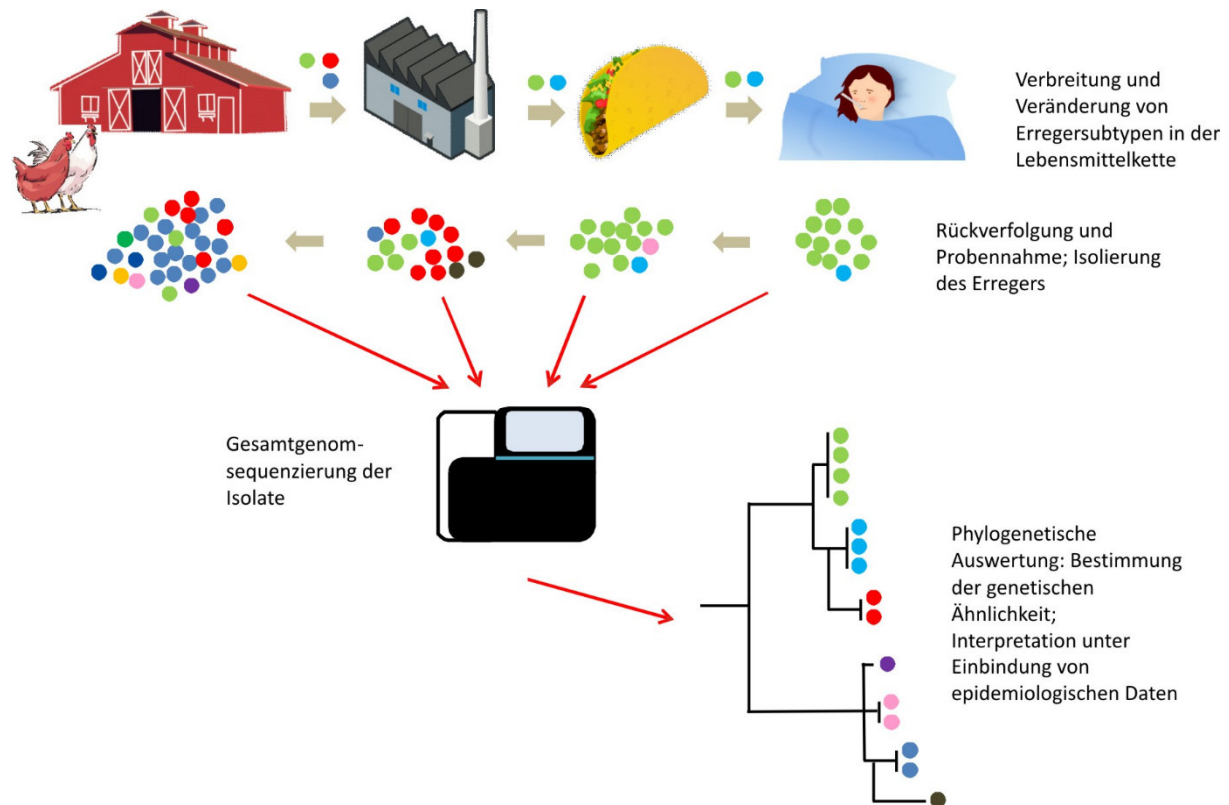
#### 3.1 Genetische Variabilität bakterieller Isolate

Lebensmittelbedingte Erreger können in kurzer Zeit in der Lebensmittelkette oder im Menschen viele Generationen durchlaufen und unterliegen somit einer ständigen Veränderung (Cowley et al. 2016). In Abhängigkeit vieler extrinsischer und intrinsischer Faktoren erfolgen diese Veränderungen schneller oder langsamer. Bietet eine genetische Veränderung dem Erreger einen Überlebensvorteil (z.B. Anpassung an bestimmte Umweltbedingungen), ist die Wahrscheinlichkeit hoch, dass diese Veränderung an weitere Nachfahren weitergegeben wird und sich in der Population des Mikroorganismus manifestiert (Ferreira et al. 2014). Dabei führt nicht jede genetische Veränderung zu einem Selektionsvorteil. Studien zur Rate einer durchschnittlichen Veränderung eines Bakteriengenoms (die sog. Mutationsrate) gehen z.B. für *S. Enteritidis* durchschnittlich von 1,01 Basen pro Genom pro Jahr aus (Deng et al. 2014) und für *E. coli* mit 1,1 Basen pro Genom pro Jahr (Didelot et al. 2012). Dagegen wurde für die Spezies *Staphylococcus aureus* eine Mutationsrate von 8,4 Basen pro Genom pro Jahr ermittelt (Didelot et al. 2012) und für *Listeria monocytogenes* wiederum von 0,18 bis 0,35 Basen (Knudsen et al. 2017).

In einer WGS-Analyse wird, wie oben beschrieben, die Anzahl der SNP- bzw. Allel-Unterschiede verwendet, um phylogenetische Bäume zu konstruieren, die Informationen über die Entwicklungsgeschichte der Isolate liefern und darstellen, wie Stämme zueinander verwandt sind. In biologischer Hinsicht bedeutet eine hohe Sequenzähnlichkeit, dass Isolate einen jungen gemeinsamen Vorfahren besitzen. Im Gegensatz bedeutet eine geringe Ähnlichkeit, dass sie im besten Fall von einem älteren gemeinsamen Vorfahren abstammen (Pightling et al. 2018). Eine grundlegende Annahme in der molekularen Epidemiologie ist, dass die Phylogenie die epidemiologische Verwandtschaft widerspiegelt, d.h., klinische Isolate oder klinische und phylogenetisch eng verwandte Lebensmittel- oder Umweltisolate wahrscheinlich in einem epidemiologischen Zusammenhang stehen (Besser et al. 2018). Bei der Aufnahme oder Abgabe von ganzen DNA-Abschnitten durch Mechanismen, die als horizontaler Gentransfer oder Rekombination bekannt sind, kann sich eine Zelle so stark in ihrem Genom verändern, dass eine phylogenetische Abbildung der Evolution nicht mehr möglich ist. Dabei ist nicht davon auszugehen, dass sich Bakterienpopulationen unter natürlichen Bedingungen in einer Lebensmittelkette evolutionär so verhalten, wie es unter Laborbedingungen beobachtet wird (Deatherage et al. 2017). Bakterienpopulationen unterliegen vielmehr verschiedenen Stufen der genetischen Diversität, die sich aus verschiedenen Vorgängen wie genetische Drift, natürliche Selektion oder Gründereffekte ergeben. Solche Ereignisse können an jedem Punkt in der Lebensmittelkette stattfinden und zu einer verminderten oder erhöhten Diversität in einer Bakterienpopulation führen (Deatherage et al. 2017).

Der Einsatz von Desinfektionsmitteln in einem Produktionsbetrieb oder der Einsatz von Antibiotika in der Primärproduktion vermindert eine Bakterienpopulation stark oder kann sie nahezu auslöschen (Jagadeesan et al. 2019). Überlebende Bakterien können jedoch eine neue Population mit neu erworbenen Eigenschaften gründen. In einem phylogenetischen Baum würde man die enge Verwandtschaft der alten und neuen Population aufgrund zahlreicher akkumulierter genetischer Veränderungen nicht notwendigerweise sofort erkennen (Colijn

and Gardy 2014). Phylogenetische Bäume reflektieren daher nicht immer direkt Übertragungsereignisse, können aber Auskunft darüber geben, wie und mit welcher Dynamik der Weg der Übertragung erfolgte (Besser 2018; Colijn and Gardy 2014; Didelot et al. 2012; Jagadeesan et al. 2019). Abbildung 2 zeigt die Entstehung, Verbreitung und Selektion eines Erregers in der Lebensmittelkette und wie sich Veränderungen bei einer Rückverfolgung der Infektionsquelle bzw. des -vehikels in einem phylogenetischen Baum widerspiegeln können.



**Abb. 2:** Entstehung, Verbreitung und Selektion eines bakteriellen Erregers in der Lebensmittelkette und Darstellung der Veränderungen in einem phylogenetischen Baum. In die Lebensmittelkette (Primärproduktion) eingetretene Erreger unterliegen einer Veränderung während der Produktion und eine Vielzahl von Varianten (dargestellt in Kreisen mit verschiedener Farbe) entsteht. Ein Teil der Varianten wird in lebensmittelverarbeitende Betriebe eingeschleppt. Dort kann es zu einer weiteren Diversifizierung der Population mit Entstehung neuer Varianten kommen. Ein Teil der Varianten wird durch den Verzehr von produzierten Lebensmitteln auf den Menschen übertragen, der sich schließlich mit einer oder mehreren im Lebensmittel vorkommenden Varianten infizieren kann. Durch eine Rückverfolgung werden aus humanen Proben, Lebensmitteln, Lebensmittelbetrieben und der Primärproduktion (einschließlich Futtermittel) Isolate gewonnen und sequenziert. Die genetische Ähnlichkeit der Sequenzen wird bioinformatisch ermittelt. Die Interpretation der Daten erfolgt unter Einbindung von epidemiologischen Daten.

### 3.2 Schwellenwerte für die Interpretation von WGS-Daten

Erfahrungswerte bei der Auswertung von WGS-Daten, die im Zusammenhang mit lebensmittelbedingten Krankheitsausbrüchen gewonnen wurden, lassen grobe Einschätzungen zu, in wie vielen SNPs oder Allelen sich Isolate unterscheiden können (Tab. 3). Wie wahrscheinlich es ist, dass zwei Isolate miteinander in einem Zusammenhang stehen, ohne zunächst weitergehende epidemiologische Informationen zu haben, untersuchten Pightling et al. (2018) anhand der Anzahl von SNPs, der Topologie des phylogenetischen Baumes und den damit verbundenen Bootstrap-Werten. Die Autoren definieren drei Bereiche, die ein Ergebnis (Match) unterstützend, neutral oder nicht unterstützend einordnen. Danach sind zwei Erregerisolate, die einen Unterschied von nicht mehr als 20 SNPs besitzen, als eng verwandt ein-

zustufen. Sie besitzen dann wahrscheinlich einen jüngeren gemeinsamen Vorfahren, der aus einer gemeinsamen Quelle abstammt. Wenn solche eng verwandten Isolate von verschiedenen Orten in einer Lebensmittelproduktionsanlage isoliert werden, ist das wahrscheinlichste Szenario, dass sich der gleiche Stamm innerhalb der Produktionsumgebung ausgebreitet hat.

Isolate mit einem SNP-Unterschied zwischen 20 und 100 werden als neutral eingestuft. In diesem Fall ist es umso wichtiger, epidemiologische Daten hinzuzuziehen, um eine richtige Schlussfolgerung ziehen zu können. Neben der Anzahl der SNPs ist aber auch der Bootstrap-Wert entscheidend, der die Robustheit eines Zweiges in einem phylogenetischen Baum widerspiegelt. Der Wert sollte hier nicht unter 0,8 liegen, wenn die Isolate verdächtigt werden, in Zusammenhang mit einem Ausbruch zu stehen. Wenn die Sequenzen zweier Isolate sehr unterschiedlich sind, z.B. mehr als 100 SNPs, gelten sie im Allgemeinen als nicht mehr verwandt. Die Wahrscheinlichkeit, dass sie nicht aus derselben Quelle stammen, ist dann sehr hoch (Pightling et al. 2018).

Die Einteilung der Isolate liegt nicht immer innerhalb der oben genannten Schwellenwerte. Beispielsweise können sich Isolate in einer Lebensmittelverarbeitungsanlage in 30 SNPs unterscheiden, aber im Vergleich zu anderen Isolaten zum selben Cluster gehören. Dies deutet darauf hin, dass die Isolate einen gemeinsamen Vorfahren aufweisen und sich wahrscheinlich von einem resilienten Stamm entwickelt haben, der in der Anlage über längere Zeit persistieren konnte (Elson et al. 2019). Dies kann wie oben beschrieben eintreten, wenn mikrobielle Populationen eine starke Verminderung ihrer Anzahl z.B. durch Desinfektionsmittel erfahren, da zufällige Mutationen zu einer Diversifizierung des Vorgängerstammes führen können (Jagadeesan et al. 2019). Die o.g. Schwellenwerte können jedoch auch bei Ausbrüchen, die mit einer Quelle verbunden sind, überschritten werden. In einem Fall waren bei der Exposition mit Salmonellen durch kleine Schildkröten gleichzeitig drei *Salmonella*-Serovare (*S. Poona*, *S. Pomona* und *S. Sandiego*) beteiligt. Die Unterschiede der assoziierten Isolate für *S. Poona* betragen bis zu 17 SNPs und für *S. Pomona* bis zu 30 SNPs (<https://www.cdc.gov/salmonella/small-turtles-03-12/epi.html>). Ebenso zeigten 401 Isolate, die in Zusammenhang mit einem multinationalen europäischen Ausbruch von *S. Enteritidis* Phagentyp 14b mit Eiern als Quelle standen, einen maximalen Unterschied von 23 SNPs (Dallman et al. 2016). Auch bei *S. Typhimurium*-Ausbrüchen ist ein Überschreiten der o.g. unterstützenden SNP-Anzahl beschrieben worden (Leekitcharoenphon et al. 2014).

Eine allgemeingültige Festlegung eines Schwellenwertes für die Definition eines Ausbruchsgeschehens ist daher schwierig und hängt von weiteren guten epidemiologischen Daten im Einzelnen ab (Jagadeesan et al. 2019; Pightling et al. 2018; Schürch et al. 2018). Bei der Zuordnung von Erkrankten zu einem Ausbruchsgeschehen und der gleichzeitigen Rückverfolgung von möglichen Kontaminationsquellen von belasteten Lebensmitteln müssen die Herkunft und das Isolationsdatum der Isolate berücksichtigt werden. Isolate von gleichzeitig Erkrankten, die nachweislich dem gleichen lebensmittelbedingten Krankheitsausbruch zugeordnet werden können, weisen erwartungsgemäß eine geringe Variabilität in ihren Genomsequenzen aus, da sie von einer gemeinsamen Infektionsquelle (Lebensmittel) stammen und im Körper der Erkrankten in der Regel nur eine kurze Verweildauer bis zum Krankheitsausbruch besteht (Deng et al. 2014). Kontaminationsquellen in einem Produktionsbetrieb können jedoch schon sehr lange aktiv sein und persistierende Erreger unterliegen einer gewissen Dynamik, die zu einer höheren Diversität in ihren Bakterienpopulationen führt. Die im Rahmen von Probenahmen in Produktionsbetrieben gewonnenen Isolate decken selten alle Variationen in einer persistierenden Bakterienpopulation ab. Daher kann häufig beobachtet werden, dass Umweltisolate, die bei Rückverfolgungen in Betriebsstätten gewonnen wurden, genetisch von Isolaten, die vom ursächlichen Lebensmittel und von Patienten stammen, genetisch weiter voneinander entfernt sind (d.h. eine höhere Anzahl an SNPs bzw. Allele zueinander aufweisen).

Auch bei der Anwendung der cgMLST-Analyse werden Schwellenwerte angegeben, die hilfreich zur Erkennung von Ausbruchsklustern sein können. Zum Beispiel hat sich für die zwei meist verwendeten *L. monocytogenes*-Schemata nach Ruppitsch (Ruppitsch et al. 2015) und nach Moura (Moura et al. 2016) ein Wert von  $\leq 7$  Alleldifferenz zwischen zwei Isolaten als sinnvoll ergeben (Van Walle et al. 2018). Dabei kann die maximale Anzahl der unterschiedlichen Allele, die in einem Cluster zusammengefasst werden, sieben übersteigen (Halbedel et al. 2018). Für ein *Salmonella*-bekanntes cgMLST-Schema nach Achtman et al. (<https://enterobase.warwick.ac.uk/>) wird ein schrittweises Clustering mit verschiedenen Schwellenwerten angeboten. Die Autoren betonen jedoch, dass es nicht offensichtlich ist, welcher Schwellenwert nützlich sein kann und epidemische Ausbruchsisolate mit Werten von  $\leq 2$ , 5 oder 10 auftreten können (Zhou et al. 2019). Grundsätzlich ist die Anzahl von SNP-Unterschieden zwischen Isolatsequenzen, die zu einem Ausbruch gezählt werden, größer als die Anzahl von cgMLST-Allelunterschieden, da bei der SNP-Analyse ein größerer Genomsequenzbereich als bei der cgMLST-Analyse berücksichtigt wird (Halbedel et al. 2018; Jagadeesan et al. 2019; Pearce et al. 2018).

Tab. 3: Beispiele für die Unterschiede der SNPs bzw. Allele von Isolaten in lebensmittelbedingten Krankheitsausbrüchen

Spezies	Ausbruchsmatrix	Ausbruchsjahr	Anzahl Isolate untersucht	Bereich SNP	Bereich wgMLST (Allele)	Bereich cgMLST (Allele)	Referenz
<i>E. coli</i> O157 (STEC)	Gefrorene Rindfleischburger	2012	36	0–106	0–90	n.u.	(Rumore et al. 2018)
<i>E. coli</i> O157 (STEC)	Rindfleischbratlinge	2012	42	0–110	0–100	n.u.	(Rumore et al. 2018)
<i>E. coli</i> O157 (STEC)	Rohmilchkäse	2013	44	0–5	0–9	n.u.	(Rumore et al. 2018)
<i>E. coli</i> O121 (STEC)	Mehl	2015–2016	40	0–2	n.u.	n.u.	(Crowe et al. 2017)
<i>E. coli</i> O157 (STEC)	Rohmilch	2017	17	0–1	n.u.	n.u.	(Treacy et al. 2019)
<i>E. coli</i> O157 (STEC)	Brunnenkresse	2013	22	0–3	n.u.	n.u.	(Jenkins et al. 2015)
<i>E. coli</i> O103 (STEC)	Rohmilch	2017	17	n.u.	n.u.	0–5	(Mylius et al. 2018)
<i>L. monocytogenes</i>	Geräuchertes Schweinefleisch	2012–2016	543	n.u.	n.u.	0–4	(Kleta et al. 2017)
<i>L. monocytogenes</i>	Fleischprodukt	2016–2017	83	n.u.	n.u.	0–7	(Pietzka et al. 2019)
<i>L. monocytogenes</i>	Krabbenfleisch	2012–2013	10	0–10	n.u.	n.u.	(Elson et al. 2019)
<i>L. monocytogenes</i>	Krabbenfleisch	2007–2013	50	0–40	n.u.	n.u.	(Elson et al. 2019)
<i>L. monocytogenes</i>	Lachsprodukte	2017	21	0–9	n.u.	0–16	(Schjørring et al. 2017)
<i>L. monocytogenes</i>	Eiscreme	2013–2018	10	0–31	n.u.	n.u.	(Allard et al. 2019)
<i>L. monocytogenes</i>	Käse	2012–2013	56	0–12	n.a.	0–9	(Chen et al. 2017)
<i>S. enterica</i>	Eier (Deutschland)	2014	207	0–23	n.u.	n.u.	(Inns et al. 2015)
<i>S. enterica</i>	Eier (Spanien)	2015	136	0–20	n.u.	n.u.	(Inns et al. 2017)
<i>S. enterica</i>	Rohwurst, Schwein	2013–2014	145	0–5	n.u.	0–6	(Simon et al. 2018)
<i>S. enterica</i>	Eier (Tasmanien)	2005–2008	12	0–23	n.u.	n.u.	(Hawkey et al. 2013)
<i>S. enterica</i>	Sesampaste	2016–2017	28	n.u.	n.u.	0–2	(Meinen et al. 2019)
<i>C. jejuni</i>	Leberpastete	2012	7	n.u.	n.u.	0–15	(Lahti et al. 2017)
<i>C. jejuni</i>	Milch	2002–2003	10	n.u.	0–12	n.u.	(Revez et al. 2014)

n.a.: untersucht, aber nicht angegeben, n.u.: nicht untersucht



## 4 Harmonisierung und Standardisierung

Um WGS-Daten laborübergreifend analysieren und vergleichen zu können, muss die Herstellung der Isolatsequenzen, deren bioinformatische Auswertung sowie der Datenaustausch Standards unterworfen sein, die eindeutige Qualitätskriterien an die erzeugten Genomsequenzen stellen. Zudem müssen auch einheitliche Auswertungsverfahren festgelegt werden. Insbesondere Auswertungsprogramme für die Clustererkennung mit der Methode der cgMLST- bzw. SNP-Analyse offerieren eine Vielzahl von Optionen, die einen direkten Vergleich von WGS-Ergebnissen zwischen den Laboren erschweren (Lüth et al. 2018). Derzeit müssen Vergleichsanalysen von Isolatsequenzen daher noch lokal an einer Stelle erfolgen und können nicht dezentral durchgeführt werden. Auf ISO-Ebene gibt es inzwischen eine Arbeitsgruppe (ISO TC 34/SC 9/WG25), die international einen Standard zur Anwendung des WGS für die Typisierung und genomische Charakterisierung von Erregern, die über das Lebensmittel übertragen werden können, erarbeitet. Ein erster Arbeitsentwurf befindet sich derzeit im Komitee Status der ISO (ISO 2019). Dieser Arbeitsentwurf umfasst Laborverfahren zur Herstellung der Sequenzen, die bioinformatische Analyse, Metadatenhinterlegung in Datenbanken und die Validierung und Verifizierung von WGS-Arbeitsabläufen. Der offizielle Prozess bis zum endgültigen Standard wird jedoch noch einige Jahre in Anspruch nehmen.

Inzwischen konnten einige Aktivitäten auf internationaler Ebene durchgeführt werden, die in Richtung Harmonisierung von WGS-Methoden weisen. Sowohl das Global-Microbial-Identifizier(GMI)-Konsortium (<https://www.globalmicrobialidentifier.org>) als auch das GenomeTrakr-Konsortium (<https://www.fda.gov/food/whole-genome-sequencing-wgs-program/genome-trakr-network>) führten eine Reihe von WGS-Ringversuchen durch, mit denen die teilnehmenden Labore die Vergleichbarkeit ihrer WGS-Ergebnisse überprüfen konnten (Moran-Gilad et al. 2015; Timme et al. 2018) (<https://www.globalmicrobialidentifier.org/workgroups/gmi-proficiency-test-reports>). Auch wurden Referenzgenomsätze veröffentlicht, mit denen die Leistungsfähigkeit von Softwaretools verglichen werden kann (Timme et al. 2017; Timme et al. 2019).

Um in Deutschland die Anwendung von WGS-Methoden zu standardisieren, hat das BVL zu Beginn dieses Jahres für den Bereich der Lebensmittelsicherheit die Arbeitsgruppe „NGS-Bakteriencharakterisierung“ im Rahmen der § 64-LFGB-Methodensammlung gegründet. Diese Gruppe besteht aus Vertretern der Labore der Länder und des Bundes sowie aus Vertretern von Privatlaboren. In der ersten Sitzung, die im März 2019 stattfand, konnte sich darauf geeinigt werden, dass zunächst Verfahren validiert werden, die für die Typisierung von bakteriellen Isolaten im Rahmen von Ausbruchsuntersuchungen angewendet werden (Szabo et al. 2019). Die Ausarbeitung von validierten Protokollen für die Genomsequenzierung und Analyse der WGS-Daten wird die Labore der Länder und des Bundes in die Lage versetzen, ihre Daten auch dezentral analysieren zu können.





## 5 Austausch von WGS-Daten

Behörden, die eine effektivere Überwachung von Erregern im Rahmen der Aufklärung von lebensmittelbedingten Krankheitsausbrüchen erreichen, und Unternehmen, die Kontaminationen im Produktionsprozess ihrer Lebensmittel nachverfolgen möchten, müssen sich mit dem Thema des WGS-Datenaustausches beschäftigen. Erreger können sich durch den globalen Welthandel und Personenreiseverkehr sehr schnell weltweit verbreiten. Die Aufklärung von Ausbrüchen ist daher auch immer mit der Notwendigkeit des Austausches von WGS-Daten und der zugehörigen epidemiologischen Daten verbunden, um eine räumliche und zeitliche Verbreitung von Erregern und Erreger-Subtypen erkennen zu können (FAO 2016). Behörden auf föderaler, nationaler, europäischer und schließlich internationaler Ebene, die in der Überwachung von lebensmittelbedingten Krankheitserregern involviert sind, benötigen daher abgestimmte Konzepte, wie ein Datenaustausch erfolgen kann (Lüth et al. 2018). Der Austausch der Daten muss dabei nicht nur innerhalb eines Sektors, sondern nach dem One-Health-Ansatz übergreifend, d.h. zwischen den für Lebensmittelsicherheit, öffentliches Gesundheitswesen, Veterinärwesen und Umwelt zuständigen Behörden, erfolgen.

Um den Datenaustausch so effizient wie möglich zu gestalten, wäre die Veröffentlichung aller WGS-Daten zu den Isolaten in einer gemeinsamen Datenbank sinnvoll. Aus rechtlichen Gründen werden sensible Metadaten zu der Genomsequenz des Isolats nicht veröffentlicht und die WGS-Daten werden deshalb lediglich mit einer Basisinformation zum Isolat in öffentliche Datenbanken eingestellt. Die Gründe hierfür sind vielfältig. Insbesondere der Missbrauch oder die Fehlinterpretation der Daten werden angeführt (Aarestrup and Koopmans 2016; Lüth et al. 2018). Dieses stellt vor allem die Hersteller eines belasteten Lebensmittels vor ein großes Dilemma: Einerseits können Krankheitsausbrüche durch den Datenaustausch schneller erkannt und gestoppt werden, andererseits müssen die Hersteller rechtliche und wirtschaftliche Konsequenzen befürchten. Dieses Problem kann nur durch den offenen Dialog zwischen den verschiedenen Sektoren unter Einbeziehung der Lebensmittelunternehmer gelöst werden. Notwendig ist eine klare Regelung zum Datenaustausch und den Handlungsabläufen, die bei einer Datenübereinstimmung erfolgen müssen. International agierende Lebensmittelkonzerne sehen bereits die Vorteile der Implementierung des WGS in ihre Qualitätssichernden Maßnahmen und verfügen schon teilweise über große Mengen von WGS-Daten im Rahmen ihrer Eigenkontrollen. Diese Konzerne sind aber auch auf ihre Reputation und den Schutz vor regulatorischen Maßnahmen der Behörden bedacht und daher häufig noch zurückhaltend, wenn es um die Veröffentlichung ihrer WGS-Daten geht (Jagadeesan et al. 2019).

Die Lebensmittelunternehmen und die zuständigen Behörden könnten durch die Bereitstellung von WGS-Daten aus Eigenkontrollen oder Anlasskontrollen schneller Infektionsquellen ausfindig machen, um zukünftig schneller reagieren zu können und damit die Lebensmittelsicherheit zu verbessern. Es zeigte sich in der Vergangenheit, dass ein großer wirtschaftlicher Schaden für einen Lebensmittelunternehmer immer dann entstanden ist, wenn der betroffene Erreger schon sehr lange – teilweise über Jahre – im Betrieb persistierte und sich unkontrolliert in der Produktionskette auf eine Vielzahl von Produkten verbreitet und somit zu vielen Erkrankungsfällen geführt hat. Solch ein Risiko wird minimiert, wenn eine schnelle Erkennung von Hygieneproblemen in einem Betrieb wahrscheinlicher wird. Unter Einbeziehung von WGS-Daten könnten Lebensmittelunternehmer ihre Hygienekonzepte anpassen und bisher unerkannte Wege einer Verschleppung von Erregern im Betriebsablauf aufklären und stoppen. Der Lebensmittelunternehmer kann unter Umständen Kontaminationseinträge verringern bzw. abstellen und dadurch größeren Schaden in der Zukunft abwenden. Grundsätzlich besteht auf behördlicher sowie auf unternehmerischer Seite der Wunsch, den WGS-Datenaustausch auf globaler Ebene zu regeln und zu vereinfachen und regulatorische Maßnahmen für die Lebensmittelindustrie nachvollziehbarer zu machen (FAO 2016).

Für die Überwachung und die Untersuchung von überregionalen Krankheitsausbrüchen in Deutschland wird ein zukunftsfähiges Konzept zum WGS-Datenmanagement für die Bundesländer benötigt. Als Ansatz ist es denkbar, die WGS-Daten der Isolate in öffentliche Datenbanken und die dazugehörigen epidemiologischen Metadaten in nichtöffentliche Datenbanken zu stellen. WGS-Daten ohne Metadaten sind zunächst wenig aussagekräftig, da eine Interpretation bzw. Rückverfolgung auf den Ursprung des Isolates nicht möglich ist. Bei einem Verdacht auf ein Ausbruchsgeschehen kann die Rückverfolgung daher nur durch die zuständigen Behörden der Länder erfolgen. In Deutschland gibt es viele klein- und mittelständische Lebensmittelunternehmen, die sich eine Qualitätssicherung zur Betriebshygiene auf der Basis von WGS-Daten nicht leisten werden können. Die zuständigen Behörden, die in den Bundesländern für die Einhaltung der Hygienebestimmungen verantwortlich sind, sollten diesbezüglich die Lebensmittelunternehmen beratend unterstützen. In einem zweiten Schritt ist zu überlegen, wie Lebensmittelunternehmen an einer prospektiven Überwachung beteiligt werden können. Hier würde es eine nochmalige Verbesserung zum frühen Auffinden von potenziell belasteten Quellen mit Erregern geben, welche ökonomische Schäden für das Unternehmen, aber auch das Risiko von gesundheitlichen Schäden für den Verbraucher verringern kann.

## 6 Schlussfolgerung

Das WGS bietet im Vergleich zu konventionellen Typisierungsmethoden ein wesentlich detaillierteres Bild von verdächtigen Isolaten, die im Rahmen von lebensmittelbedingten Krankheitsausbrüchen untersucht werden. Mit dem WGS ist die Echtzeit-Überwachung von Erregern und Erreger-Subtypen möglich. Daher stellt das WGS ein sehr effektives Tool für die zuständigen Behörden dar, um lebensmittelbedingte Krankheitsausbrüche früh zu erkennen und geeignete Maßnahmen zu ergreifen. Diese Technologie muss jedoch noch weiter standardisiert und harmonisiert werden. Gelingt dies auf nationaler und schließlich internationaler Ebene und werden Übereinkommen abgeschlossen, Daten (WGS-Daten und epidemiologische Daten), die zur Aufklärung von Ausbrüchen benötigt werden, zwischen den zuständigen Behörden im Bereich der öffentlichen Gesundheit, Veterinärmedizin, Lebensmittelsicherheit und Umwelt auszutauschen und im Idealfall auch Daten von Lebensmittelproduzenten mit einzuschließen, wird dies höchstwahrscheinlich zu geringeren Fallzahlen von lebensmittelbedingten Infektionserkrankungen führen und somit einen hohen gesundheitlichen und ökonomischen Nutzen darstellen. Wir verweisen zu diesem Thema auch auf die kürzlich veröffentlichte Scientific Opinion der EFSA, abrufbar unter: <https://efsa.onlinelibrary.wiley.com/doi/10.2903/j.efsa.2019.5898> (EFSA BIOHAZ Panel et al. 2019).



## **7 Anmerkungen**

Das Bundesinstitut für Risikobewertung hat für Laura Uelze eine finanzielle Förderung durch den Bund aufgrund eines Beschlusses des Deutschen Bundestages durch die Bundesregierung im Rahmen des Projektes „Integrierte genombasierte Surveillance von Salmonellen (GenoSalmSurv), Zuweisungsbescheid ZMVI1-2518FSB709 vom 26.11.2018, erhalten. Josephine Grützke wurde unterstützt durch eine finanzielle Förderung des Ministeriums für Bildung und Forschung im Rahmen des Drittmittelprojektes „Lebensmittelsicherheit und Resilienz von Lebensmittelwarenketten in biologischen Gefahrenlagen“ (Ess-B.A.R.), FKZ 13N13982.

### **Erklärung**

Die Autoren erklären, dass sie keine Interessenkonflikte haben.



## 8 Referenzen

- Aarestrup FM, Koopmans MG (2016) Sharing data for global infectious disease surveillance and outbreak detection. *Trends Microbiol* 24:241–241, doi: 10.1016/j.tim.2016.01.009
- Aigrain L, Gu Y, Quail MA (2016) Quantitation of next generation sequencing library preparation protocol efficiencies using droplet digital PCR assays – a systematic comparison of DNA library preparation kits for Illumina sequencing. *BMC Genomics* 17:458 doi:10.1186/s12864-016-2757-4
- Allard MW, Strain E, Melka D, Bunning K, Musser SM, Brown EW, Timme R (2016) Practical value of food pathogen traceability through building a whole-genome sequencing network and database. *J Clin Microbiol* 54:1975–1983 doi:10.1128/JCM.00081-16
- Allard MW et al. (2019) Whole genome sequencing uses for foodborne contamination and compliance: discovery of an emerging contamination event in an ice cream facility using whole genome sequencing. *Infect Genet Evol* 73:214–220 doi:10.1016/j.meegid.2019.04.026
- Ashton PM et al. (2016) Identification of *Salmonella* for public health surveillance using whole genome sequencing. *PeerJ* 4:e1752 doi:10.7717/peerj.1752
- Becker L, Steglich M, Fuchs S, Werner G, Nubel U (2016) Comparison of six commercial kits to extract bacterial chromosome and plasmid DNA for MiSeq sequencing. *Sci Rep* 6:28063 doi:10.1038/srep28063
- Besser J, Carleton HA, Gerner-Smidt P, Lindsey RL, Trees E (2018) Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clin Microbiol Infect* 24:335–341 doi:10.1016/j.cmi.2017.10.013
- Besser JM (2018) *Salmonella* epidemiology: A whirlwind of change *Food Microbiol* 71:55–59 doi:10.1016/j.fm.2017.08.018
- Bielaszewska M et al. (2011) Characterisation of the *Escherichia coli* strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. *Lancet Infect Dis* 11:671–676 doi:10.1016/S1473-3099(11)70165-7
- Bogaerts B et al. (2019) Validation of a bioinformatics workflow for routine analysis of Whole-Genome Sequencing data and related challenges for pathogen typing in a European National Reference Center: *Neisseria meningitidis* as a proof-of-concept. *Front Microbiol* 10:362 doi:10.3389/fmicb.2019.00362
- Cao Y, Fanning S, Proos S, Jordan K, Srikumar S (2017) A Review on the applications of Next Generation Sequencing technologies as applied to food-related microbiome studies. *Front Microbiol* 8:1829 doi:10.3389/fmicb.2017.01829
- Carrico JA, Rossi M, Moran-Gilad J, Van Domselaar G, Ramirez M (2018) A primer on microbial bioinformatics for nonbioinformaticians. *Clin Microbiol Infect* 24:342–349 doi:10.1016/j.cmi.2017.12.015
- Chen Y et al. (2017) Whole Genome and Core Genome Multilocus Sequence Typing and Single Nucleotide Polymorphism analyses of *Listeria monocytogenes* isolates associated with an outbreak linked to cheese, United States, 2013. *Appl Environ Microbiol* 83 doi:10.1128/AEM.00633-17
- Colijn C, Gardy J (2014) Phylogenetic tree shapes resolve disease transmission patterns. *Evol Med Public Health* 2014:96-108 doi:10.1093/emph/eou018
- Cowley LA et al. (2016) Short-term evolution of Shiga toxin-producing *Escherichia coli* O157:H7 between two food-borne outbreaks. *Microb Genom* 2:e000084 doi:10.1099/mgen.0.000084
- Crowe SJ et al. (2017) Shiga toxin-producing *E. coli* infections associated with flour. *N Engl J Med* 377:2036–2043 doi:10.1056/NEJMoa1615910
- Dallman T et al. (2016) Phylogenetic structure of European *Salmonella* Enteritidis outbreak correlates with national and international egg distribution network. *Microb Genom* 2:e000070 doi:10.1099/mgen.0.000070
- de Lannoy C, de Ridder D, Risse J (2017) The long reads ahead: de novo genome assembly using the MinION F1000. *Res* 6:1083 doi:10.12688/f1000research.12012.2

- Deatherage DE, Kepner JL, Bennett AF, Lenski RE, Barrick JE (2017) Specificity of genome evolution in experimental populations of *Escherichia coli* evolved at different temperatures. Proc Natl Acad Sci U S A 114:E1904–E1912 doi:10.1073/pnas.1616132114
- Deng X et al. (2014) Genomic epidemiology of *Salmonella enterica* serotype Enteritidis based on population structure of prevalent lineages. Emerg Infect Dis 20:1481–1489 doi:10.3201/eid2009.131095
- Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW (2012) Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet 13:601–612 doi:10.1038/nrg3226
- ECDC et al. (2019) EFSA and ECDC technical report on the collection and analysis of whole genome sequencing data from food-borne pathogens and other relevant microorganisms isolated of from human, animal, food, feed and food/feed environmental samples in the joint ECDC-EFSA molecular typing database. EFSA supporting publication 2019 EN-1337:92 pp doi:10.2903/sp.efsa.2019.EN-1337
- EFSA BIOHAZ Panel et al. (2019) Whole genome sequencing and metagenomics for outbreak investigation, source attribution and risk assessment of food-borne microorganisms. EFSA Journal 17:79 doi:10.2903/j.efsa.2019.5898
- Elson R et al. (2019) Utility of whole genome sequencing to describe the persistence and evolution of *Listeria monocytogenes* strains within crabmeat processing environments linked to two outbreaks of listeriosis. J Food Prot 82:30–38 doi:10.4315/0362-028X.JFP-18-206
- Endrullat C, Glokler J, Franke P, Frohme M (2016) Standardization and quality management in next-generation sequencing. Appl Transl Genom 10:2–9 doi:10.1016/j.atg.2016.06.001
- FAO (2016) Applications of whole genome sequencing in food safety management. Verfügbar unter: <http://www.fao.org/3/a-i5619e.pdf>,
- Ferreira V, Wiedmann M, Teixeira P, Stasiewicz MJ (2014) *Listeria monocytogenes* persistence in food-associated environments: epidemiology, strain characteristics, and implications for public health. J Food Prot 77:150–170 doi:10.4315/0362-028X.JFP-13-150
- Gardy JL, Loman NJ (2018) Towards a genomics-informed, real-time, global pathogen surveillance system. Nat Rev Genet 19:9–20 doi:10.1038/nrg.2017.88
- Halbedel S, Prager R, Fuchs S, Trost E, Werner G, Flieger A (2018) Whole-Genome Sequencing of recent *Listeria monocytogenes* isolates from Germany reveals population structure and disease clusters. J Clin Microbiol 56 doi:10.1128/JCM.00119-18
- Hawkey J, Edwards DJ, Dimovski K, Hiley L, Billman-Jacobe H, Hogg G, Holt KE (2013) Evidence of microevolution of *Salmonella* Typhimurium during a series of egg-associated outbreaks linked to a single chicken farm. BMC Genomics 14:800 doi:10.1186/1471-2164-14-800
- Heather JM, Chain B (2016) The sequence of sequencers: The history of sequencing DNA. Genomics 107:1–8 doi:10.1016/j.ygeno.2015.11.003
- Hendriksen R et al. (2018) Final report of ENGAGE – Establishing Next Generation sequencing ability for genomic analysis in Europe. Annex D-List of online bioinformatics tools and software used for capacity building (status January 2018). EFSA supporting publication 2018 EN-1431:58-78
- Hutchins RJ, Phan KL, Saboor A, Miller JD, Muehlenbachs A, Workgroup CNQ (2019) Practical guidance to implementing quality management systems in public health laboratories performing Next-Generation Sequencing: personnel, equipment, and process management (phase 1). J Clin Microbiol 57 doi:10.1128/JCM.00261-19
- Inns T et al. (2017) Prospective use of whole genome sequencing (WGS) detected a multi-country outbreak of *Salmonella* Enteritidis. Epidemiol Infect 145:289–298 doi:10.1017/S0950268816001941



- Inns T et al. (2015) A multi-country *Salmonella* Enteritidis phage type 14b outbreak associated with eggs from a German producer: 'near real-time' application of whole genome sequencing and food chain investigations, United Kingdom, May to September 2014. *Euro Surveill* 20 doi:10.2807/1560-7917.es2015.20.16.21098
- ISO (2019) Microbiology of the food chain — Whole genome sequencing for typing and genomic characterization of foodborne bacteria — General requirements and guidance (ISO/CD 23418[E]). ISO, Geneva, Switzerland
- Jackson BR et al. (2016) Implementation of nationwide real-time whole-genome sequencing to enhance listeriosis outbreak detection and investigation. *Clin Infect Dis* 63:380–386 doi:10.1093/cid/ciw242
- Jagadeesan B et al. (2019) The use of next generation sequencing for improving food safety: translation into practice. *Food Microbiology* 79:96–115 doi:10.1016/j.fm.2018.11.005
- Jain S, Mukhopadhyay K, Thomassin PJ (2019) An economic analysis of salmonella detection in fresh produce, poultry, and eggs using whole genome sequencing technology in Canada. *Food Res Int* 116:802–809 doi:10.1016/j.foodres.2018.09.014
- Jenkins C, Dallman TJ, Grant KA (2019) Impact of whole genome sequencing on the investigation of food-borne outbreaks of Shiga toxin-producing *Escherichia coli* serogroup O157:H7, England, 2013 to 2017. *Euro Surveill* 24 doi:10.2807/1560-7917.ES.2019.24.4.1800346
- Jenkins C et al. (2015) Public health investigation of two outbreaks of Shiga toxin-producing *Escherichia coli* O157 associated with consumption of watercress. *Appl Environ Microbiol* 81:3946–3952 doi:10.1128/AEM.04188-14
- Kleta S et al. (2017) Molecular tracing to find source of protracted invasive listeriosis outbreak, Southern Germany, 2012–2016. *Emerg Infect Dis* 23:1680-1683 doi:10.3201/eid2310.161623
- Knudsen GM, Nielsen JB, Marvig RL, Ng Y, Worning P, Westh H, Gram L (2017) Genome-wide-analyses of *Listeria monocytogenes* from food-processing plants reveal clonal diversity and date the emergence of persisting sequence types. *Environ Microbiol Rep* 9:428–440 doi:10.1111/1758-2229.12552
- Kovac J, den Bakker H, Carroll LM, Wiedmann M (2017) Precision food safety: A systems approach to food safety facilitated by genomics tools Trac-Trend. *Anal Chem* 96:52–61 doi:10.1016/j.trac.2017.06.001
- Kwong JC, McCallum N, Sintchenko V, Howden BP (2015) Whole genome sequencing in clinical and public health microbiology. *Pathology* 47:199–210 doi:10.1097/Pat.0000000000000235
- Lahti E, Lofdahl M, Agren J, Hansson I, Olsson Engvall E (2017) Confirmation of a campylobacteriosis outbreak associated with chicken liver paté using PFGE and WGS. *Zoonoses Public Health* 64:14–20 doi:10.1111/zph.12272
- Leekitcharoenphon P, Nielsen EM, Kaas RS, Lund O, Aarestrup FM (2014) Evaluation of whole genome sequencing for outbreak detection of *Salmonella enterica*. *PLoS One* 9:e87991 doi:10.1371/journal.pone.0087991
- Loman NJ et al. (2012a) High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol* 10:599–606 doi:10.1038/nrmicro2850
- Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ (2012b) Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 30:434–439 doi:10.1038/nbt.2198
- Luheshi L et al. (2015) Pathogen genomics into practice. PHG Foundation, Cambridge, UK
- Lüth S, Kleta S, Al Dahouk S (2018) Whole genome sequencing as a typing tool for foodborne pathogens like *Listeria monocytogenes* – The way towards global harmonisation and data exchange. *Trends Food Sci Tech* 73:67–75 doi:10.1016/j.tifs.2018.01.008
- Maiden MC (2006) Multilocus sequence typing of bacteria. *Annu Rev Microbiol* 60:561–588 doi:10.1146/annurev.micro.59.030804.121325

- Maiden MC, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND (2013) MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol* 11:728–736 doi:10.1038/nrmicro3093
- Meinen A et al. (2019) Salmonellosis outbreak with novel *Salmonella enterica* subspecies *enterica* serotype 11:z41:e,n,z15 attributable to sesame products in five european countries, 2016 to 2017. *Euro Surveill* 24 doi:10.2807/1560–7917 ES.2019.24.36.1800543
- Moran-Gilad J et al. (2015) Proficiency testing for bacterial whole genome sequencing: an end-user survey of current capabilities, requirements and priorities. *BMC Infect Dis* 15:174 doi:10.1186/s12879-015-0902-3
- Moura A et al. (2016) Whole genome-based population biology and epidemiological surveillance of *Listeria monocytogenes*. *Nat Microbiol* 2:16185 doi:10.1038/nmicrobiol.2016.185
- Mylius M et al. (2018) Shiga toxin-producing *Escherichia coli* O103:H2 outbreak in Germany after school trip to Austria due to raw cow milk, 2017 – The important role of international collaboration for outbreak investigations. *Int J Med Microbiol* 308:539-544, doi:101016/j.ijmm.2018.05.005
- Nadon C et al. (2017) PulseNet International: Vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. *Euro Surveill* 22 doi:10.2807/1560-7917.ES.2017.22.23.30544
- Pallen MJ (2013) Reply to Updating benchtop sequencing performance comparison. *Nat Biotechnol* 31:296 doi:10.1038/nbt.2531
- Pearce ME, Alikhan NF, Dallman TJ, Zhou Z, Grant K, Maiden MCJ (2018) Comparative analysis of core genome MLST and SNP typing within a European *Salmonella* serovar Enteritidis outbreak. *Int J Food Microbiol* 274:1–11 doi:10.1016/j.ijfoodmicro.2018.02.023
- Pérez-Losada M, Arenas M, Castro-Nallar E (2018) Microbial sequence typing in the genomic era. *Infect Genet Evol* 63:346–359 doi:10.1016/j.meegid.2017.09.022
- Pietzka A et al. (2019) Whole genome sequencing based surveillance of *L. monocytogenes* for early detection and investigations of listeriosis outbreaks. *Front Public Health* 7:139 doi:10.3389/fpubh.2019.00139
- Pightling AW, Pettengill JB, Luo Y, Baugher JD, Rand H, Strain E (2018) Interpreting whole-genome sequence analyses of foodborne bacteria for regulatory applications and outbreak investigations. *Front Microbiol* 9:1482 doi:10.3389/fmicb.2018.01482
- Quainoo S, Coolen JPM, van Hijum S, Huynen MA, Melchers WJG, van Schaik W, Wertheim HFL (2017) Whole-Genome Sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. *Clin Microbiol Rev* 30:1015–1063 doi:10.1128/CMR.00016-17
- Revez J, Zhang J, Schott T, Kivisto R, Rossi M, Hanninen ML (2014) Genomic variation between *Campylobacter jejuni* isolates associated with milk-borne-disease outbreaks. *J Clin Microbiol* 52:2782–2786 doi:10.1128/JCM.00931-14
- RKI (2018) Infektionsepidemiologisches Jahrbuch meldepflichtiger Krankheiten für 2017. Robert Koch-Institut, Berlin. doi:10.17886/rkipubl-2018-001
- Ronholm J, Nasheri N, Petronella N, Pagotto F (2016) Navigating microbiological food safety in the era of whole-genome sequencing. *Clin Microbiol Rev* 29:837–857 doi:10.1128/Cmr.00056-16
- Rosner B, Mikolajetz U, Schonsky A (2018) Gemeinsamer nationaler Bericht des BVL und RKI zu lebensmittelbedingten Krankheitsausbrüchen in Deutschland 2017. Bundesamt für Verbraucherschutz und Lebensmittelsicherheit (BVL), Berlin, Deutschland
- Rumore J et al. (2018) Evaluation of whole-genome sequencing for outbreak detection of Verotoxigenic *Escherichia coli* O157:H7 from the Canadian perspective. *BMC Genomics* 19:870 doi:10.1186/s12864-018-5243-3

- Ruppitsch W et al. (2015) Defining and evaluating a core genome Multilocus Sequence Typing scheme for Whole-Genome Sequence-based typing of *Listeria monocytogenes*. J Clin Microbiol 53:2869–2876 doi:10.1128/JCM.01193-15
- Sanger F, Coulson AR (1975) Rapid method for determining sequences in DNA by primed synthesis with DNA-polymerase. J Mol Biol 94:441–+ doi:10.1016/0022-2836(75)90213-2
- Sato MP et al. (2019) Comparison of the sequencing bias of currently available library preparation kits for Illumina sequencing of bacterial genomes and metagenomes. DNA Res 26:391–398 doi:10.1093/dnares/dsz017
- Schjørring S et al. (2017) Cross-border outbreak of listeriosis caused by cold-smoked salmon, revealed by integrated surveillance and whole genome sequencing (WGS), Denmark and France, 2015 to 2017. Euro Surveill 22 doi:10.2807/1560-7917.ES.2017.22.50.17-00762
- Schürch AC, Arredondo-Alonso S, Willems RJL, Goering RV (2018) Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. Clin Microbiol Infect 24:350–354 doi:10.1016/j.cmi.2017.12.016
- Sekse C, Holst-Jensen A, Dobrindt U, Johannessen GS, Li W, Spilsberg B, Shi J (2017) High throughput sequencing for detection of foodborne pathogens. Front Microbiol 8:2029 doi:10.3389/fmicb.2017.02029
- Seth-Smith HMB, Bonfiglio F, Cuenod A, Reist J, Egli A, Wuthrich D (2019) Evaluation of rapid library preparation protocols for Whole Genome Sequencing based outbreak investigation. Front Public Health 7:241 doi:10.3389/fpubh.2019.00241
- Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, Waterston RH (2017) DNA sequencing at 40: past, present and future. Nature 550:345–353 doi:10.1038/nature24286
- Simon S et al. (2018) Evaluation of WGS based approaches for investigating a food-borne outbreak caused by *Salmonella enterica* serovar Derby in Germany. Food Microbiol 71:46–54 doi:10.1016/j.fm.2017.08.017
- Stevens EL, Timme R, Brown EW, Allard MW, Strain E, Bunning K, Musser S (2017) The Public health impact of a publically available, environmental database of microbial genomes. Front Microbiol 8:808 doi:10.3389/fmicb.2017.00808
- Szabo K, Malorny B, Stoyke M (2019) Etablierung der § 64 LFGB Arbeitsgruppen „NGS – Bakteriencharakterisierung“ und „NGS – Speziesidentifizierung“. J Verbrauch Lebensm, veröffentlicht am 22. Oktober 2019. doi.org/10.1007/s00003-019-01255-z
- Taboada EN, Graham MR, Carrico JA, Van Domselaar G (2017) Food safety in the age of next generation sequencing, bioinformatics, and open data access. Front Microbiol 8:909 doi:10.3389/fmicb.2017.00909
- Timme RE et al. (2018) GenomeTrakr proficiency testing for foodborne pathogen surveillance: an exercise from 2015. Microb Genom 4 doi:10.1099/mgen.0.000185
- Timme RE et al. (2017) Benchmark datasets for phylogenomic pipeline validation, applications for foodborne pathogen surveillance. PeerJ 5:e3893 doi:10.7717/peerj.3893
- Timme RE et al. (2019) Phylogenomic pipeline validation for foodborne pathogen disease surveillance. J Clin Microbiol 57 doi:10.1128/JCM.01816-18
- Treacy J et al. (2019) Outbreak of Shiga toxin-producing *Escherichia coli* O157:H7 linked to raw drinking milk resolved by rapid application of advanced pathogen characterisation methods, England, August to October 2017. Euro Surveill 24 doi:10.2807/1560-7917.ES.2019.24.16.1800191
- Tyler AD et al. (2016) Comparison of sample preparation methods used for the next-generation sequencing of *Mycobacterium tuberculosis*. PLoS One 11:e0148676 doi:10.1371/journal.pone.0148676
- Van Walle I et al. (2018) Retrospective validation of whole genome sequencing-enhanced surveillance of listeriosis in Europe, 2010 to 2015. Euro Surveill 23 doi:10.2807/1560-7917.ES.2018.23.33.1700798

- Weiser AA, Thons C, Filter M, Falenski A, Appel B, Käsbohrer A (2016) FoodChain-Lab: A trace-back and trace-forward tool developed and applied during food-borne disease outbreak investigations in Germany and Europe. PLoS One 11:e0151977 doi:10.1371/journal.pone.0151977
- WHO (2018) Whole genome sequencing for foodborne disease surveillance. Landscape paper. Verfügbar unter: [https://www.who.int/foodsafety/publications/foodborne\\_disease/wgs\\_landscape/en/](https://www.who.int/foodsafety/publications/foodborne_disease/wgs_landscape/en/),
- Wilson DJ (2012) Insights from genomics into bacterial pathogen populations. PLoS Pathog 8:e1002874 doi:10.1371/journal.ppat.1002874
- Zähringer H (2014) Die DNA Schnipselmacher. Produktübersicht: Kits zur Herstellung von NGS-Bibliotheken. Laborjournal 9
- Zhou Z, Alikhan N-F, Mohamed K, Fan Y, Agama Study Group, Achtman M (2020) The Enterobase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny and *Escherichia* core genomic diversity. Genome Res 30: 138-152, doi:10.1101/gr.251678.119

## 9 Abbildungsverzeichnis

- Abb. 1: Methoden der WGS-Datenauswertung im Rahmen von Krankheitsausbrüchen zur Bestimmung der genetischen Ähnlichkeit von Isolatsequenzen. (A) Gene-by-Gene Annäherung (Allelbestimmung von Genen), (B) Referenzmapping (SNP-Bestimmung anhand eines Referenzgenoms). 10
- Abb. 2: Entstehung, Verbreitung und Selektion eines bakteriellen Erregers in der Lebensmittelkette und Darstellung der Veränderungen in einem phylogenetischen Baum. In die Lebensmittelkette (Primärproduktion) eingetretene Erreger unterliegen einer Veränderung während der Produktion und eine Vielzahl von Varianten (dargestellt in Kreisen mit verschiedener Farbe) entsteht. Ein Teil der Varianten wird in lebensmittelverarbeitende Betriebe eingeschleppt. Dort kann es zu einer weiteren Diversifizierung der Population mit Entstehung neuer Varianten kommen. Ein Teil der Varianten wird durch den Verzehr von produzierten Lebensmitteln auf den Menschen übertragen, der sich schließlich mit einer oder mehreren im Lebensmittel vorkommenden Varianten infizieren kann. Durch eine Rückverfolgung werden aus humanen Proben, Lebensmitteln, Lebensmittelbetrieben und der Primärproduktion (einschließlich Futtermittel) Isolate gewonnen und sequenziert. Die genetische Ähnlichkeit der Sequenzen wird bioinformatisch ermittelt. Die Interpretation der Daten erfolgt unter Einbindung von epidemiologischen Daten. 18



## 10 Tabellenverzeichnis

Tab. 1: Benötigte Infrastruktur für ein molekular- bzw. mikrobiologisch arbeitendes Labor zur Genomsequenzierung von bakteriellen Isolaten mittels WGS im Rahmen der Aufklärung von Krankheitsausbrüchen	11
Tab. 2: Mindestens anzustrebende Qualitätsparameter der Sequenzierung für ausgesuchte über das Lebensmittel übertragbare Erreger	13
Tab. 3: Beispiele für die Unterschiede der SNPs bzw. Allele von Isolaten in lebensmittelbedingten Krankheitsausbrüchen	21





## 11 Glossar

**Abdeckungstiefe:** gibt Aufschluss darüber, wie häufig jede Base des Genoms durchschnittlich sequenziert wurde.

**Bibliothek:** Sammlung von genomischen DNA-Fragmenten eines Isolats, die für die Erstellung der Genomsequenz des Isolats verwendet wird.

**Bibliothekenherstellung:** experimenteller Prozess zur Herstellung einer DNA-Bibliothek eines Isolats für die NGS. Für die Herstellung wird die aus dem Isolat aufgereinigte DNA zunächst enzymatisch oder mechanisch fragmentiert. An die Enden der fragmentierten DNA werden Adaptersequenzen, die für den Sequenzierungsprozess benötigt werden, ligiert. Anschließend erfolgen eine Amplifizierung der Fragmente mit der PCR und ein Aufreinigungsschritt.

**Bioinformatik:** Sammlung, Speicherung und Auswertung von biologischen Sequenzdaten mit informationstechnischen Hilfsmitteln.

**Contig:** Zusammenhängender Abschnitt einer DNA-Sequenz, der sich aus der Überlappung kürzerer DNA-Sequenzen ergibt.

**„De-novo“-Assemblierung:** Prozess der Anordnung von kurzen DNA-Sequenzen zu längeren DNA-Sequenzen (Contigs). Viele kürzere Sequenzen werden zu längeren Sequenzstücken zusammengesetzt.

**Epidemiologische Daten:** bezeichnet einen Datensatz, der die Probeneinheit (z.B. Datum und Ort der Probenahme, Art der Probe und Herkunft der Probe, z.B. Tier/Lebensmittel/Futtermittel) beschreibt. Wenn aus der Probe ein Bakterienisolat erhalten werden konnte, welches molekularbiologisch (z.B. mit WGS) typisiert wird, muss eine Verlinkung dieser Daten mit den molekularen Typisierungsdaten erfolgen.

**Gene-by-Gene-Annäherung:** Verfahren, bei dem Sequenzen an ausgewählte Gensequenzen angeordnet werden. Unterschiede in der Sequenz werden durch eine Allelnummer, die für die Gensequenz spezifisch ist, beschrieben.

**Metadaten:** Daten, die andere Daten definieren und beschreiben. Metadaten können der Probenahme, dem Isolat oder der Sequenz zugeordnet werden. Metadaten sollten je nach Probentyp (epidemiologische Daten), entsprechend den durchgeführten Tests oder durchgeführten Operationen bereitgestellt werden, die die gespeicherten Sequenzierungsdaten beschreiben.

**MLST:** bezeichnet die Sequenzierung mehrerer Gene oder eines genetischen Locus, der genügend polymorphe Stellen aufweist, um in einem Typisierungsschema verwendet zu werden. Bei der cgMLST werden mehrere hundert solcher Gene berücksichtigt, die üblicherweise in allen Isolaten der betrachteten Spezies vorhanden sind („Kerngene“). Bei der wgMLST werden alle Gene einer Spezies in der Analyse berücksichtigt, unabhängig von der Abwesenheit/Anwesenheit der Gene in allen Isolaten.

**Next Generation Sequencing:** Hochdurchsatzverfahren zum Bestimmen der Nukleotidsequenz eines Genoms oder eines Teils davon. Diese Technik verwendet DNA-Sequenzierungstechnologien, die in der Lage sind, mehrere DNA-Sequenzen parallel zu verarbeiten. Auch als massiv parallele Sequenzierung und NGS bezeichnet.

**Paired-end-Sequenzierung:** Prozess, der die Sequenzierung beider Enden eines DNA-Fragmentes ermöglicht und so den Nachweis repetitiver Sequenzelemente erleichtert.

**Pipeline/bioinformatische Auswertungspipeline:** Programme, Skripte oder Software-Module, die aufeinander aufbauen. Das Ergebnis eines Programms wird als Input für eine nachfolgende Datenverarbeitung verwendet.

**Phylogenie:** Abbildung der evolutionären Beziehung von Organismen.

**Referenzbasiertes Mapping:** Prozess, bei dem Sequenzen an eine vorgegebene Referenzsequenz angeordnet werden. Dies erfolgt mithilfe von Software.

**Sequenztyp (ST):** numerische Bezeichnung für ein bestimmtes allelisches DNA-Sequenzprofil. Ursprünglich sind sieben Loci indiziert, für die jeder einzelnen Sequenz eine beliebige und eindeutige Allelnummer zugeordnet ist, die in das allelische Profil integriert ist. STs werden in Sequenz-Typisierungsschemata als Vergleichseinheit verwendet, die auf allelischen Varianten basiert. Isolate, die für alle Sequenzen identische Allele besitzen, werden einem gemeinsamen Sequenztyp (ST) zugeordnet.

**SNP:** Ein Single-Nukleotid-Polymorphismus (SNP) ist eine Substitution eines einzelnen Nukleotids, das an einer bestimmten Stelle im Genom auftritt.

**Trimming:** Adaptersequenzen werden von den Rohsequenzen entfernt und Basen mit schlechter Qualität aus der Sequenz herausgefiltert.

**Whole Genome Sequencing:** Verfahren zur Bestimmung der DNA-Sequenz eines Genoms unter Verwendung der gesamten genomischen DNA einer Probe.