

DOI 10.17590/20191128-102235

Erbmaterial von Erregern vergleichen, um Krankheitsausbrüche aufzuklären

Stellungnahme Nr. 047/2019 des BfR vom 28. November 2019

Lebensmittel können mit krankmachenden Erregern wie Bakterien, Viren oder Parasiten verunreinigt sein. Durch den weltweiten Handel von Lebensmittelprodukten können Verunreinigungen an weit voneinander entfernten Orten zu Krankheitsausbrüchen führen. Um Ursache und mögliche Zusammenhänge von Krankheitsgeschehen zeitnah aufzuklären, nutzen die zuständigen Kontrollbehörden molekulare Labormethoden. Diese werden durch die Anwendung auch permanent weiterentwickelt.

Zu dieser Thematik hat sich das Bundesinstitut für Risikobewertung (BfR) mit relevanten Fragen befasst und bewertet, inwieweit molekulare Methoden der Next Generation Sequencing-Verfahren zur Aufklärung von Krankheitsausbruchsgeschehen geeignet und welche Daten darüber hinaus einzubeziehen sind. Die vorliegende Bewertung dient vor allem den Behörden im Bereich der öffentlichen Gesundheit, Lebensmittelsicherheit und Veterinärmedizin, um zu entscheiden, welches Verfahren bei der Aufklärung von Ausbruchsgeschehen geeignet ist.

„Next Generation Sequencing (NGS)“, steht für die zweite und dritte Generation der Sequenzierung von Erbmaterial und bietet die höchst mögliche Auflösung, um die Nukleotidabfolge eines DNA-Moleküls oder Genoms zu bestimmen. Die Kosten der Methode haben sich in den letzten Jahren deutlich verringert.

Um Krankheitsausbrüche zu untersuchen kommen derzeit zwei Verfahren in Frage:

1) Für leicht kultivierbare Erreger, die in reiner Form vorliegen, hat sich weltweit die Gesamtgenomsequenzierung (engl.: whole genome sequencing, WGS) etabliert. Dabei wird die Erbsubstanz des ursächlichen Keims von Erkrankten isoliert und mit Isolaten des gleichen Erregers aus Lebensmitteln verglichen. So lassen sich kleinste Unterschiede im Erbmaterial erkennen und auswerten.

2) Bei der Gesamtmetagenomsequenzierung (engl.: whole metagenome sequencing, WMS) dagegen wird Erbsubstanz aus einer Lebensmittel-Probe direkt gewonnen, die oft verschiedene Keime enthält. Damit lassen sich auch schwer bis nicht kultivierbare Mikroorganismen wie Parasiten oder Viren nachweisen. Die Gesamtmetagenomsequenzierung eignet sich als eine Methode zur ersten Diagnostik, wenn kein Verdacht auf einen konkreten Erreger vorliegt.

Um lebensmittelbedingte Krankheitsausbrüche aufzuklären, ist es unabhängig vom angewendeten Analyseverfahren notwendig, epidemiologische Daten zu den untersuchten Erregern hinzuzuziehen. So lässt sich etwa bewerten, ob ein nachgewiesener Erreger zu einem bestimmten Ausbruchsgeschehen gehört. Im Idealfall gelingt es, Übertragungswege und die Quelle der Verunreinigung sicher aufzuklären.

1 Gegenstand der Bewertung/ Einleitung

Mikroorganismen unterliegen einer ständigen Veränderung von Merkmalen über ihre Generationen hinweg. In Abhängigkeit vieler extrinsischer und intrinsischer Faktoren erfolgt die Veränderung schneller oder langsamer. Bietet die Veränderung dem Mikroorganismus einen Überlebensvorteil (z. B. Anpassung an die Umweltbedingungen), ist die Wahrscheinlichkeit

hoch, dass sie an weitere Nachfahren weitergegeben wird und sich in der Population des Mikroorganismus manifestiert. Die Veränderung kann jedoch auch einen Nachteil für die Zelle bringen, sodass sie sich nicht weiter verbreiten kann. Die so im Laufe der Evolution erworbenen neuen Eigenschaften sind nicht immer sichtbar. Ihre Veränderung ist jedoch in der Nukleotidsequenz des Genoms hinterlegt. Studien zur Rate einer durchschnittlichen Veränderung eines Bakteriengenoms, die sog. Mutationsrate, gehen z. B. für *Escherichia coli* von ca. 1×10^{-3} Mutationen pro Genom pro Generation aus (Lee et al., 2012). Für das *Salmonella* Serovar *S. Choleraesuis* var. Kunzendorf konnte die Mutationsrate in Bezug auf die Zeit mit 1,02 Basen pro Genom pro Jahr ermittelt werden (Leekitcharoenphon et al. 2019). Die Rate für *S. Enteritidis* wurde in einer Studie ähnlich mit 1,01 Basen pro Jahr ermittelt (Deng et al., 2014).

Molekulare Typisierungsmethoden sind wichtige Werkzeuge, um Unterschiede im genetischen Material eines Isolates durch den Vergleich mit weiteren Isolaten der gleichen Spezies zu erkennen. Das Aufspüren solcher Unterschiede in einer Bakterienpopulation, die in kurzen Zeiträumen (wenige Jahre) stattgefunden haben, benötigt daher hoch diskriminierende Typisierungsmethoden. Im Rahmen von Krankheitsausbruchsgeschehen verbreitet sich ein sog. Ausbruchsstamm in der Regel durch direkten Kontakt von Menschen oder durch die Aufnahme des Keims über belastete Lebensmittel sehr schnell. Im ungünstigen Falle erfolgt die Verbreitung über größere Gebiete in Form von vielen Ausbruchsklustern. Ein weltweiter Handel von Lebensmitteln fördert eine solche Verbreitung. In der Vergangenheit wurden in der Ausbruchsauflösung für einen Vergleich verdächtiger zusammenhängender Isolate die Multilocus variable-number tandem-repeats Analyse (MLVA) und die Pulsfeldgelelektrophorese (PFGE) eingesetzt. Allerdings haben diese Methoden nur eine begrenzte Auflösungskraft, sodass häufig eine Kombination von Typisierungsmethoden in der Ausbruchsauflösung verwendet wurde.

Unter dem Begriff „DNA-Sequenzierung“ versteht man die Bestimmung der Nukleotidabfolge (Sequenz) in einem DNA-Molekül. Das erste Verfahren, womit man eine solche Sequenz bestimmen konnte, wurde von Sanger Ende der 70er Jahre entwickelt. Hiermit konnten in einer begrenzten Zahl kurze DNA Abschnitte sequenziert werden. Diese Methode ist jedoch nicht geeignet, in kurzer Zeit und kosteneffizient Genome vollständig sequenzieren zu können. Mit zunehmender Bedeutung der Sequenzierung in der Forschung und Diagnostik kam Mitte der 2000er die nächste Generation (zweite Generation) der Sequenzierertechnologie auf den Markt. Mit den dafür entwickelten Geräten konnte eine massenhafte parallele Sequenzierung von kurzen Nukleotidabfolgen eines DNA Moleküls durchgeführt werden. Unter Verwendung von bioinformatischen Ansätzen ermöglichen sie es, die Nukleotidabfolge ganzer Genome bestimmen zu können (z. B. *E. coli* Genom von 5 Megabasen). Inzwischen ist eine weitere Sequenzierertechnologie (Dritte-Generation) verfügbar, die einzelne Moleküle sequenzieren kann. Dadurch ist der Schritt der Polymerase-Kettenreaktion, der noch in der zweiten Generation eingesetzt wurde, nicht mehr notwendig. Sequenzierer der dritten Generation ermöglichen es, sehr lange DNA-Moleküle (ca. 10-50 Kilobasen) am Stück sequenzieren zu können (Ronholm et al., 2016).

Das „Next Generation Sequencing (NGS)“ (zweite und dritte Generation Sequencing) bietet die höchst mögliche Auflösungskraft für die Bestimmung der Nukleotidabfolge eines DNA-Moleküls oder Genoms. Die Kosten des NGS haben sich in den letzten Jahren soweit verringert, dass der Einsatz der Technologie für Behörden im Bereich der öffentlichen Gesundheit, Veterinärmedizin und Lebensmittelsicherheit erschwinglich geworden ist. Dies führte weltweit zum Einsatz des NGS, um die Verwandtschaft von Isolaten im Rahmen von Untersuchungen zu Krankheitsausbrüchen oder der Überwachung und des Monitoring von Kontaminationen in der Lebensmittelproduktion zu bestimmen. Die Kombination dieser Sequenzvergleiche mit

relevanten epidemiologischen Informationen und idealerweise mit der Auswertung der Warketteninformationen erlaubt es schließlich, die Übertragungswege und Quelle sicher aufzuklären. Der Erfolg einer Ausbruchsaufklärung erfordert daher die kontinuierliche Zusammenarbeit der Mikrobiologen und Epidemiologen, um die Bewertung genomischer und epidemiologischer Beweise zu einem sicheren Ergebnis zusammenzufügen (World Health Organisation, 2018).

Zu dieser Thematik hat sich das Bundesinstitut für Risikobewertung (BfR) mit folgenden Fragen befasst:

1. NGS-Verfahren

- a.) *Eignen sich alle NGS-Verfahren in gleicher Weise für die Untersuchung im Rahmen von Krankheitsausbrüchen?*
- b.) *Welche Vor- und Nachteile der einzelnen technischen Varianten sind für die Untersuchung von Probenmaterial im Rahmen von Krankheitsausbruchsgeschehen von Bedeutung?*
- c.) *Welchen qualitätssichernden Maßnahmen sind die angewandten NGS-Verfahren (z. B. im Rahmen von Ringversuchen) unterzogen worden?*

2. Analyseergebnisse

- a.) *Welcher Übereinstimmungsgrad ist bei Ergebnissen von Untersuchungen mit NGS-Verfahren erforderlich, um sie einem bestimmten Krankheitsausbruchsgeschehen, z. B. sog. Cluster zuzuordnen?*
- b.) *Wie hoch ist der Aussagewert der verschiedenen Techniken der NGS-Verfahren?*
- c.) *Welche Informationen sind erforderlich, um ein Cluster / einen identischen molekularbiologischen Subtyp einheitlich zu charakterisieren?*

3. Daten zum Untersuchungsmaterial

- a.) *Welche betriebs- oder produktbezogenen Angaben werden im Zusammenhang mit übersandten Lebensmittelisolaten und Typisierungsergebnissen (aus Probenmaterial der für die Lebensmittelüberwachung zuständigen Länder) zu welchem Zweck erfragt (oder erhoben) und wie werden diese zugeordnet?*
- b.) *Wie und von welcher Stelle sollen diese Angaben im Falle eines Ausbruchsgeschehens genutzt werden?*

Die Fragen werden wie folgt beantwortet:

1. NGS-Verfahren:

- a.) *Eignen sich alle NGS-Verfahren in gleicher Weise für die Untersuchung im Rahmen von Krankheitsausbrüchen?*

NGS-Verfahren erlauben die massenhafte Bestimmung von Nukletidabfolgen von genetischem Material in einer Probe. Sie setzen sich in der Regel (i) aus der Gewinnung der Nukleinsäure aus einer Probe, (ii) der Bibliothekenherstellung, (iii) der massenhaften Sequenzierung der Nukleinsäure und (iv) der bioinformatischen Auswertung der Rohsequenzierungsstücke zusammen. Grundsätzlich wird zwischen DNA oder RNA, die aus einer Reinkultur (z. B. Isolat) oder aus einer komplexen Zusammensetzung von verschiedenen Organismen (z. B. einer Lebensmittelprobe) gewonnen wurde, unterschieden.

Für die Untersuchung von Krankheitsausbrüchen kommen derzeit zwei Verfahren in Frage, deren Einsatz jedoch abhängig von der Ausgangslage des Geschehens ist. Das bei leicht kultivierbaren Erregern angewandte Verfahren ist die Gesamtgenomsequenzierung (engl.:

whole genome sequencing (WGS)). Hierbei wird das Genom sequenziert, das von einem in Reinform vorliegenden Bakterienisolat stammt und anschließend mit weiteren ebenso sequenzierten Genomen verdächtiger Isolate verglichen wird. Durch diesen Ansatz können einzelne Nukleotidunterschiede zwischen den Genomen über größere Sequenzabschnitte festgestellt werden. Aufgrund seiner Genauigkeit und hohen Auflösungskraft hat sich WGS weltweit für die Untersuchung von relevanten und gut kultivierbaren Erregern im Rahmen der Aufklärung von Krankheitsausbrüchen etabliert.

Ein weiterer Ansatz für die Untersuchung im Rahmen von Krankheitsausbrüchen ist die Gesamtmetagenomsequenzierung (engl.: whole metagenome sequencing (WMS)). Dies ist ein Verfahren bei dem die DNA aus einer Probe, wie z. B. aus einem Lebensmittel, extrahiert wird und diese im Anschluss mittels NGS analysiert wird. Hierfür wird auf die häufig zeitaufwendige Isolation und Anzucht der Erreger verzichtet. Dies ermöglicht, neben dem zeitlichen Vorteil, auch eine Detektion von schwer bis nicht-kultivierbaren Mikroorganismen wie Parasiten oder Viren. Da bei der WMS keine Anzucht der Erreger vor ihrer Sequenzierung vorgenommen wird, kann jedoch keine Aussage über die Lebens- und Proliferationsfähigkeit der Erreger getroffen werden. Weiterhin erfordert das Verfahren aufwendige bioinformatische Analysen, da die komplexen Sequenzgemische den einzelnen Erregern zugeordnet werden müssen. Eine geringe Sequenzabdeckung kann einen Vergleich zwischen den Proben außerdem erschweren. Das Verfahren eignet sich daher als eine universelle und ungerichtete Methode bei Abwesenheit eines spezifischen Verdachtsfalls zur Vorfelddiagnostik und nur unter bestimmten Voraussetzungen zur Untersuchung von Krankheitsausbrüchen.

b.) Welche Vor- und Nachteile der einzelnen technischen Varianten sind für die Untersuchung von Probenmaterial im Rahmen von Krankheitsausbruchsgeschehen von Bedeutung?

Wie oben beschrieben, gibt es unter den NGS-Technologien solche der zweiten und solche der dritten Generation. Für die Isolatsequenzierung von ausbruchsrelevanten Proben kommt insbesondere die zweite Generation der Sequenzierung zum Einsatz, da diese einen extrem hohen Durchsatz sowie äußerst geringe Sequenzfehlerquoten hat. Zur Detektion von Ausbruchsklustern werden Sequenzunterschiede zwischen verschiedenen Isolaten bestimmt. Daher ist es entscheidend, dass mögliche technisch bedingte Sequenzierfehler von echten biologischen Signalen unterschieden werden können. Dies ist nur möglich, wenn Sequenzierfehler zufällig und ohne erkennbares Muster auftreten, da sie so relativ zuverlässig anhand statistischer Verfahren aufgelöst werden können. Die einzige Technologie, die keine systematischen Fehler generiert, ist aktuell das „sequencing-by-synthesis“ Verfahren (Cao et al., 2017). Eine weitere Sequenzieretechnologie, das sogenannte „semi-conductor sequencing“ Verfahren generiert systematische Fehlerprofile, die einen direkten Vergleich zu den Ergebnissen aus Sequenzierungen mit dem „sequencing-by-synthesis“ Verfahren erschweren. Da sich das „sequencing-by-synthesis“ Verfahren weltweit in der Genomsequenzierung weitgehend durchgesetzt hat, können Sequenzergebnisse für die Untersuchung von Ausbruchsgeschehen mit diesem Verfahren zuverlässig verglichen werden.

Kürzlich hat sich die § 64 Lebensmittel- und Futtermittelgesetzbuch (LFGB) Arbeitsgruppe „NGS Bakteriencharakterisierung“ etabliert. Sie wird vom Bundesamt für Verbraucherschutz und Lebensmittelsicherheit (BVL) koordiniert. Ziel ist es, den mit der Lebensmittelüberwachung betrauten Behörden validierte, leistungsfähige und standardisierte Methoden der NGS-Verfahren für bakterielle Erreger lebensmittelbedingter Ausbruchsuntersuchungen zur Verfügung zu stellen. Eine Aktivität dieser Arbeitsgruppe wird die Vergleichbarkeit der verschiedenen Zweite-Generation Sequenzierungsverfahren bei der Anwendung von Ausbruchsuntersuchungen sein.

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 dem „sequencing-by-synthesis“ Verfahren zwischen 2x 75 bp bis 2x 300 bp und mit dem „semi-conductor“ Verfahren zwischen 200 bis 400 bp. Einzubeziehen in den Prozess der Sequenzierung sind aber auch die o. g. weiteren Vorbereitungsschritte der Bibliothekenherstellung. Die Erfahrungen des BfR mit dem „sequencing-by-synthesis“ Verfahren zeigen, dass die Bearbeitungsdauer bei einer Sequenzierung von kürzeren Sequenzstücken (2x 75 bp) von der vorliegenden DNA bis zum ausgewerteten Ergebnis ca. 2 Tage benötigt. Die Erzeugung von längeren Sequenzen, die eine Auswirkung auf die Anzahl der Contigs (Anzahl der bioinformatisch zusammengesetzten Sequenzstücke) haben kann, bedingt ca. 4 Tage Bearbeitungszeit. Die Anzucht des Erregers für die DNA Isolierung ist in dieser Rechnung nicht einbezogen und kann je nach Erreger weitere 2 bis 4 Tage erfordern.

Die Dritte-Generation der Sequenzierung ist in der Lage wenige, aber dafür äußerst lange Sequenzen zu generieren und kann so genutzt werden, um Referenzgenome für einzelne Ausbruchsgeschehen zu erstellen. Die Fehlerraten der zur dritten Generation gehörenden Technologien sind noch vergleichsweise hoch, so dass diese noch nicht für Isolatsequenzvergleiche im Rahmen von Untersuchungen bei Krankheitsausbrüchen geeignet sind. Sie sind jedoch sehr nützlich, um geschlossene Referenzisolatsequenzen zu erzeugen, die in Kombination mit Sequenzdaten aus der zweiten Generation kombiniert werden können und so eine sehr genaue Genomsequenz ergeben. Referenzsequenzen von Isolaten sind für die bioinformatische Auswertung notwendig. Für die meisten relevanten Erreger steht daher auch schon eine Vielzahl guter Referenzsequenzen zur Verfügung.

c.) Welchen qualitätssichernden Maßnahmen sind die angewandten NGS-Verfahren (z. B. im Rahmen von Ringversuchen) unterzogen worden?

Derzeit sind NGS Verfahren am BfR aufgrund ihrer Komplexität noch nicht akkreditiert. Es werden aber bereits verschiedene qualitätssichernde Maßnahmen während der Bibliothekenherstellung, des Sequenzierungsprozesses auf dem Gerät und der bioinformatischen Auswertung durchgeführt.

Zu den qualitätssichernden Maßnahmen während der Bibliothekenherstellung gehören sowohl die Konzentrationsüberprüfung der DNA als auch eine DNA-Fragmentlängenbestimmung. Diese Daten sind eine wichtige Voraussetzung, dass die Sequenzierung auf dem Gerät qualitativ hochwertig durchgeführt werden kann. Bei der anschließenden bioinformatischen Auswertung werden die erzeugten Rohdaten für jedes sequenzierte Isolat getrimmt (Basen mit schlechter Qualität werden herausgefiltert) und assembliert (viele kürzere Sequenzen werden bioinformatisch zu längeren Sequenzstücken zusammengesetzt). Die Qualitätsprüfung erfolgt schließlich anhand eines automatisch erstellten Assembly Reports, der wesentliche Qualitätsparameter auflistet, die üblicherweise benutzt werden, um die Qualität der Sequenzdaten einschätzen zu können. Geprüft wird die anhand der Daten vorhergesagte Spezies, die Datenmenge bzw. die Abdeckungstiefe (Coverage) des Assemblies (gibt Aufschluss darüber, wie häufig jedes Nukleotid des Genoms durchschnittlich sequenziert wurde), die Anzahl der Sequenzstücke (Contigs) sowie die Länge des Gesamt-Assemblies, die der erwarteten Länge des Genomes entsprechen sollte (gibt Aufschluss über eventuelle Kontaminationen). Die dabei zur Anwendung kommenden Grenzwerte, anhand derer bestimmt wird, ob die Qualität der Sequenzdaten ausreichend ist, basieren einerseits auf eigener Erfahrung, andererseits auf dem Austausch mit Wissenschaftlern des National Food Institute der Technical University of Denmark (DTU-Food) und der U. S. Food and Drug Administration (FDA), sowie Literaturangaben.

Jede größere (z. B. Wechsel der Sequenzierungskits) oder kleinere (z. B. Verkürzung der Sequenzierungszyklenzahl) Anpassung im Ablaufschema wird sowohl im Labor als auch bioinformatisch validiert bzw. verifiziert. Dabei kommen Kontroll-Isolate zum Einsatz, anhand derer abgeschätzt wird, ob sich bioinformatische Kennwerte (Coverage, Contigzahl, Assembly, Länge) über das akzeptable Maß hinaus verändern, bzw. ob Sequenzunterschiede mit dem ursprünglich angewandten Protokoll beobachtet werden.

Zur weiteren Qualitätssicherung der praktischen Laborarbeit, aber auch der bioinformatischen Analysen, nimmt das BfR regelmäßig an WGS Ringversuchen teil. Dazu gehörten bisher folgende Leistungsvergleichstests:

- Global Microbial Identifier Initiative (Teilnahme 2015, 2016 und 2017)
- Verschiedene Ringversuche im Rahmen des ENGAGE (Establishing Next Generation sequencing Ability for Genomic analysis in Europe) und des COMPARE (Collaborative management platform for detection and analyses of (re-) emerging and food-borne outbreaks in Europe) Projektes
- Sequenzierungsvergleichsversuche von definierten Isolaten mit dem RKI
- Bioinformatischer Ringversuch UNSGM (Practical Exercise in Support of the United Nations Secretary-General's Mechanism for Investigation of Alleged Use of Biological Weapons, with Special Consideration of the Functional Subunit Approach) zur Detektion von Pathogenitätsfaktoren in Sequenzdaten

Das BfR strebt die Akkreditierung der NGS-Verfahren für die Sequenzierung von Erregerisolaten im Rahmen von Krankheitsausbrüchen an. Vorbereitende Maßnahmen haben diesbezüglich begonnen. Das BfR hat in der kürzlich vom BVL neu eingerichteten § 64 LFGB-Arbeitsgruppe „NGS – Bakteriencharakterisierung“ auch die Obmann-Funktion übernommen und bringt seine bisher gewonnene Expertise aktiv in den Standardisierungsprozess ein. Dafür wird es Ringversuche mit den Mitgliedern der Arbeitsgruppe organisieren und durchführen. § 64 LFGB-Arbeitsgruppen werden vom BVL eingerichtet und dienen dazu, den mit der Lebensmittelüberwachung betrauten Behörden validierte, leistungsfähige und standardisierte Methoden im Rahmen der „Amtlichen Sammlung von Verfahren zur Probenahme und Untersuchung von Lebensmitteln“ (ASU) zur Verfügung zu stellen.

2. Analyseergebnisse:

Das BfR bezieht sich bei der Beantwortung der Fragen auf die verschiedenen bioinformatischen Analysemöglichkeiten, die in der Ausbruchuntersuchung von Isolatsequenzen mittels WGS eingesetzt werden. Es wird zum besseren Verständnis mit der Beantwortung von Frage 2b begonnen.

b.) Wie hoch ist der Aussagewert der verschiedenen Techniken der NGS-Verfahren?

Derzeit werden mehrere Ansätze zur Analyse von WGS-Ergebnissen in der Ausbruchuntersuchung verwendet: (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 hochwertiger Single-Nucleotide Polymorphism (SNP)-basierter Clustering-Pipelines (Kovac et al., 2017). CgMLST/wgMLST beruhen auf dem MLST-Konzept. Hier wird die DNA-Sequenzvariation in einem Set von festgelegten Genen bestimmt. Diese sind auf dem Genom verteilt und weisen trotz der Konservierung bestimmte Variationen aufgrund von Mutationen oder Rekombinationen auf. Die

Typisierung der Isolate erfolgt anhand ihres spezifischen Allelprofils der Gene, das sich schließlich in ihrem entsprechenden Sequenztyp (ST) widerspiegelt. Während die MLST auf einer „*de novo* assembly“ basierten Analyse stattfindet (Gen für Gen Annäherung) basiert die SNP-Analyse auf einem Vergleich der Einzelpunktmutationen der Isolatsequenzen zu einem vorher festgelegten, nahverwandten Referenzgenom. Diese Art der Analyse nennt man „Referenz-basiertes Mapping“. In diesem Verfahren werden keine zu analysierenden Gene festgelegt. So können auch SNPs außerhalb von genkodierenden Sequenzen berücksichtigt werden.

Während bei SNP-basierten Clustering-Verfahren alle Unterschiede in einem Set berücksichtigt werden, die zum Referenzgenom auftreten, 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. Der große Vorteil von cgMLST/wgMLST ist die Anwendung von Nomenklaturschemata, die es ermöglichen, eine eindeutige Identifizierung der zu untersuchenden Sequenz zu erstellen, die leicht kommuniziert werden kann, ohne Rohsequenzen teilen zu müssen. CgMLST/wgMLST erfordern auch weniger Aufwand bei der Rechenleistung im Vergleich zu SNP-basierten Verfahren. Durch die verschiedenen bioinformatischen Möglichkeiten, die Qualität der SNPs zu bewerten und daraus schließlich eine Auswahl 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 MLST-Gruppe. Für beide Verfahren (SNP – cgMLST/wgMLST) gilt grundsätzlich, dass Ergebnisse nur dann vergleichbar sind, wenn sie mit den gleichen oder äquivalenten Softwareprogrammen bzw. identischen Softwarealgorithmen generiert wurden. Bei SNP-Analysen müssen zudem die exakt gleichen Referenzen benutzt werden, da es sonst zu einem unterschiedlichen Ergebnis der gefundenen SNPs kommen kann. Für cgMLST-Analysen hingegen muss einerseits ein einheitliches Schema verwendet werden und außerdem eine zentral verwaltete Nomenklatur für neue Allelnummern verwendet werden. Letzteres ist angestrebt, aber derzeit noch nicht zentral vorhanden. Bisher liegen nur verschiedene isolierte Lösungen vor. Für eine korrekte Bewertung von cgMLST/wgMLST ist daher eine zentrale Datenanalyse durch Austausch der Sequenzierungsrohdaten geboten.

Die Aussagekraft von cgMLST/wgMLST als auch von SNP-basierten Analysen ist grundsätzlich für die Bestimmung der Unterschiede zwischen Isolaten im Rahmen von Ausbruchuntersuchungen sehr gut geeignet und führt zu ähnlichen Ergebnissen. Beide erlauben die Einteilung von Gruppen von Genomen in Subtypen (Clustertypen) und beide erlauben eine Bewertung, ob zwei Genome molekularbiologisch (nahezu) identisch sind. Mögliche Schwellenwerte für die Alleldistanz bzw. SNP-Distanz sind dennoch prinzipiell verschieden. Die Anwendung von cgMLST ist geeigneter, wenn mehrere Benutzer gleichzeitig jedes neue Isolat, das einer gemeinsamen Datenbank hinzugefügt wird, systematisch analysieren müssen (z. B. im Ausbruchfall), insbesondere wenn die Sequenzinformationen nicht öffentlich zugänglich sind. Für die Untersuchung der Phylogenie kann die Verwendung von cgMLST oder SNP-Verfahren robustere Analysen liefern als wgMLST, da sie nur Regionen des Genoms umfasst, die in allen Stämmen vorhanden sind. WgMLST kann aber durch die zusätzliche Berücksichtigung der variablen Gene einer Spezies zu einer höheren Auflösung führen. 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, sollten zur besseren Beurteilung beide Analyseverfahren durchgeführt werden.

Unabhängig vom Analyseverfahren ist es erforderlich, epidemiologische Daten der verdächtigen Isolate für eine Aufklärung hinzuzuziehen. Nur so können schließlich Rückschlüsse auf die Zugehörigkeit dieser zu einem Ausbruchsgeschehen oder zu einem sog. Cluster gezogen werden. Isolate, die in einem Cluster zusammengefasst werden, müssen nicht notwendigerweise zu einem Ausbruchsgeschehen gehören, wenn epidemiologische Daten zu den Sequenzvergleichsdaten im Widerspruch stehen.

a.) *Welcher Übereinstimmungsgrad ist bei Ergebnissen von Untersuchungen mit NGS-Verfahren erforderlich, um sie einem bestimmten Krankheitsausbruchsgeschehen, z. B. sog. Cluster, zuzuordnen?*

c.) *Welche Informationen sind erforderlich, um ein Cluster / einen identischen molekularen biologischen Subtyp einheitlich zu charakterisieren?*

Die Fragen 2a und 2c werden gemeinsam beantwortet, da es eine große Überlappung bei der Fragestellung gibt.

In der WGS-Analyse wird die Anzahl der SNP- bzw. Allel-Unterschiede verwendet, um phylogenetische Bäume zu konstruieren, die Informationen über die Entwicklungsgeschichte der Isolate liefern. In biologischer Hinsicht bedeutet eine hohe Sequenzähnlichkeit durch die WGS-Analyse, dass Isolate einen jungen gemeinsamen Vorfahren besitzen. Im Gegensatz bedeutet eine geringe Ähnlichkeit, dass sie im besten Fall von einem älteren gemeinsamen Vorfahren abstammen (Plichtling et al., 2018). Es ist eine grundlegende Annahme in der molekularen Epidemiologie, dass die Phylogenie die epidemiologische Verwandtschaft widerspiegelt, d. h. klinische Isolate oder klinische und phylogenetisch eng verwandte Lebensmittel- oder Umweltisolate sind wahrscheinlich epidemiologisch oder kausal verbunden (Besser et al., 2018). Diese Annahme ist aber nicht immer zutreffend, da es sich um komplexe oder indirekte Verbindungen handeln kann, die an jedem Punkt entlang der Lebensmittelkette auftreten können. Daher ist es von entscheidender Bedeutung, dass epidemiologische und lebensmittelrechtliche Rückschlüsse gemeinsam gezogen werden, um zu einer schlüssigen Interpretation der WGS-Analyse zu kommen. Die WGS-Analyse liefert belastbare Beweise dafür, dass Isolate genetisch verwandt sind, aber sie bedeutet nicht unbedingt, dass ein klinischer Fall direkt mit einem Lebensmittel in Zusammenhang steht. Es ist daher unerlässlich, dass epidemiologische Informationen vorliegen, um die phylogenetischen Ergebnisse zu unterstützen (Jagadeesan et al., 2019).

Aufgrund der Vielfalt der Bakterienarten, der unterschiedlichen epidemiologischen Kontexte und der unterschiedlichen WGS-Analyseansätze sollte die Festlegung eines Schwellenwertes, z. B. in der Interpretation während eines Ausbruchsgeschehens, sowohl bei der cgMLST als auch bei der SNP-Analyse grundsätzlich vermieden werden (Pichtling et al., 2018; Schürch et al., 2018; Jagadeesan et al., 2019). Einige Spezies oder Serotypen sind weniger divers als andere, z. B. *Salmonella* Enteritidis ist relativ klonal (Allard et al., 2013). Darüber hinaus kann die Umgebung, in der sich eine Spezies befindet, evolutionären Druck ausüben, der die Mutationsrate und die Generationszeit beeinflusst (Deatherage et al., 2017). Daher muss die Interpretation der genetischen Verwandtschaft von Stämmen, die auf SNP- oder Allelunterschieden basieren, durch Expertenwissen über den jeweiligen Krankheitserreger, einschließlich eines Verständnisses seiner genetischen Vielfalt in der Lebensmittelkette und der Repräsentativität der untersuchten Isolate, ergänzt werden (Besser et al., 2018; Schürch et al., 2018; Jagadeesan et al., 2019). Die WGS-Analyse jedes lebensmittelbedingten Ausbruchsszenarios muss unabhängig voneinander bewertet werden, wobei epidemiologische und lebensmittelkettenbezogene Untersuchungen durchgeführt werden sollten, um so viele Informationen wie möglich für die Interpretation bereitzustellen.

Für eine erste grobe Einschätzung kann man davon ausgehen, dass zwei Erregerisolate, die einen Unterschied von z. B. 0-20 SNP/Allele besitzen, als eng verwandt gelten (Jagadeesan et al, 2019). Sie besitzen dann wahrscheinlich einen jüngeren gemeinsamen Vorfahren, der aus einer gemeinsamen Quelle abstammt. Wenn solche hoch verwandten Isolate von verschiedenen Orten in einer Lebensmittelproduktionsanlage isoliert werden, ist das wahrscheinlichste Szenario, dass sich der gleiche Stamm innerhalb der Produktionsumgebung ausgebreitet hat.

Durch verschiedene Ausbruchsgeschehen wurde gezeigt, dass Isolate von erkrankten Menschen, die in einer SNP-basierten Analyse unter einen Schwellenwert von wenigen SNP fallen, eine starke zeitliche Korrelation in Bezug auf das Datum des Auftretens der Symptome aufweisen (Deng et al, 2014). Eine Untersuchung zu sieben verschiedenen *S. Enteritidis* Ausbrüchen zeigte eine Divergenz von 3 SNPs innerhalb eines Ausbruchs und die nächstgelegenen Nichtausbruchstämme unterschieden sich um durchschnittlich 42,4 SNPs (Taylor et al, 2016). Aber auch über einen längeren Ausbruchszeitraum kann die Anzahl der SNP-Unterschiede zwischen den Isolaten gering bleiben. In einer Ausbruchsstudie zu *Listeria monocytogenes* betrug die Variabilität der Isolatsequenzen innerhalb von zwei Ausbruchsklustern 5 SNPs über einen Zeitraum von drei Jahren (Gillesberg Lassen, 2016). Die Daten wurden in Kombination mit epidemiologischen Untersuchungen in den Lebensmittelbetrieben und Befragungen der Patienten bestätigt.

Wenn die Sequenzen zweier Isolate sehr unterschiedlich sind, z. B. > 50-100 SNPs oder Allele, gelten die Isolate im Allgemeinen als nicht verwandt. Die Wahrscheinlichkeit, dass sie nicht aus derselben Quelle stammen, ist dann sehr hoch.

Isolate liegen nicht immer innerhalb der oben genannten Schwellenwerte. Beispielsweise können sich Isolate in einer Lebensmittelverarbeitungsanlage in 30 SNPs/Allele unterscheiden, aber zum selben Cluster im Vergleich zu anderen Isolaten gehören. Dies deutet darauf hin, dass die Isolate einen gemeinsamen Vorfahren haben und sich wahrscheinlich von einem resilienten Stamm entwickelt haben, der in der Anlage persistiert (Elson et al., 2019). Dies kann 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>) (Jagadeesan et al, 2019). 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 SNP (Dallman et al., 2016).

3. Daten zum Untersuchungsmaterial:

a.) Welche betriebs- oder produktbezogenen Angaben werden im Zusammenhang mit übersandten Lebensmittelisolaten und Typisierungsergebnissen (aus Probenmaterial der für die Lebensmittelüberwachung zuständigen Länder) zu welchem Zweck erfragt (oder erhoben) und wie werden diese zugeordnet?

Für die Einsendung von Isolaten aus diversen Matrices (Lebensmittel, Futtermittel, Primärproduktion, Produktionsumgebung und Sonstige, ausgenommen humane Isolate) zur Untersuchung in den mikrobiologischen Nationalen Referenzlaboratorien (NRL) und Konsiliarlaboratorien im BfR steht ein elektronisches Einsendeformular im Excel-Dateiformat auf der Homepage des BfR zur Verfügung. Alle Einsender werden gebeten, das Einsendeformular möglichst vollständig ausgefüllt elektronisch per E-Mail und/oder in Papierform an das BfR zu übermitteln. Die Daten werden im Laborinformationssystem dokumentiert.

Im BfR-Einsendeformular werden folgende Metadaten in Textform und als ADV-Codes abgefragt: Probennummer, AVV-Datanummer, Erreger-Vorbefund, Datum der Probenahme, Datum der Isolation, Ort der Probenahme, Matrix, Verarbeitungszustand, Grund der Probenahme, Betriebsart, Viehverkehrsverordnungsnummer.

Lediglich für Proben aus dem Zoonosen-Monitoring sind diese Angaben Pflicht. Für alle anderen Proben müssen mindestens Probennummer, Erreger, Datum der Probenahme und Matrix an das BfR übermittelt werden.

Die Metadaten werden vom BfR für die Zuordnung einzelner Isolate eines Erregers im Rahmen

- von Ausbruchsuntersuchungen
- von epidemiologischen Fragestellungen
- der Umsetzung der Zoonosen-Überwachungsrichtlinie 2003/99/EG
- der Zoonosen-Bekämpfungsverordnung (EG) 2160/2003
- der Erhebungen von Statistiken bezüglich der Prävalenz der Erreger in unterschiedlichen Matrices außerhalb der oben genannten Angelegenheiten und Programme erhoben.

Das BfR erhält nur in Ausnahmefällen und auf freiwilliger Basis präzise Angaben zu Herstellern von Proben, welche eine eindeutige Zuordnung von Proben zu Einzelbetrieben ermöglichen. In der Regel verbleiben diese für die Ausbruchsuntersuchungen wichtigen Angaben in den Ländern und werden im Bedarfsfall vom BVL in den betroffenen Bundesländern über die obersten Landesbehörden abgefragt.

Wünschen die im Rahmen einer Ausbruchsauflärung beteiligten Landesbehörden eine Warenstromanalyse mit Rückverfolgung von verdächtigen Lebensmitteln, unterstützt das BfR auf Nachfrage sowohl bei der Datenerfassung als auch bei den Analysen, Visualisierungen und deren Bewertung mithilfe der zu diesem Zweck entwickelten Software FoodChain-Lab.

b.) Wie und von welcher Stelle sollen diese Angaben im Falle eines Ausbruchsgeschehens genutzt werden?

Um eine effiziente Untersuchung und eine sichere Interpretation der Ergebnisse von lebensmittelbedingten Krankheitsausbrüchen durchzuführen, ist es notwendig, die vorliegenden Sequenzdaten von verdächtigen Isolaten mit den epidemiologischen Daten zu verknüpfen. Sequenz- und Metadaten sollten daher in Datenbanken, die miteinander verknüpft sind, vorliegen. Die Analyse der Isolatsequenzen müssen Experten durchführen, die die Verwandtschaft der zu vergleichenden Isolate bioinformatisch ermitteln. Sind Isolate als eng verwandt eingestuft, bildet dies die Grundlage für die weitergehenden epidemiologischen Untersuchungen (z. B. Zuordnung von Proben zu Einzelbetrieben, Händlern, Maßnahmen für Kontrolluntersuchungen in den Betrieben,...).

Der Sequenzvergleich sollte durch ein Erreger-Expertenlabor durchgeführt werden. Dies kann ein Labor, das die Expertise für phylogenetische Vergleiche des jeweiligen Erregers erlangt hat, sein. Das Führen und die Verwaltung eines Metadatensatzes für eine erste epidemiologische Beurteilung sollten durch eine autorisierte Stelle erfolgen, die hierzu die notwendige Erfahrung hat (Übermittlung von Daten durch die Behörden). Diese erste Beurteilung muss keine sensitiven Daten zu Betrieben enthalten (z. B. Betriebsname oder -ort). Diese können entsprechend ihrer Zuständigkeit in den jeweiligen Bundesländern verbleiben und erst bei erhöhtem Verdacht durch das betroffene Bundesland in einer vertieften epidemiologischen Untersuchung einbezogen werden. Die Isolatsequenzdaten sollten möglichst schnell in ein öffentliches Repositorium (z. B. ENA oder NCBI) gestellt werden. Dies bietet die Grundlage für den Austausch der Sequenzdaten und die Möglichkeit weitergehende lokale Analysen unter identischen Softwarebedingungen durchführen zu können. So ist auch die Verfügung der Sequenzen bei Untersuchungen im Rahmen internationaler Ausbruchsgeschehen gesichert.

Weitere Informationen auf der BfR-Website zum Thema Next Generation Sequencing

Einsendeformular für Isolate und Proben (Excel-Dateiformat)

https://www.bfr.bund.de/de/einsendeformular_fuer_isolate_und_proben-9257.html

Forschungsprojekt *Errichtung der Next-Generation-Sequenzierung für die Genomanalyse von bakteriellen Erregern in Europa* (ENGAGE)

https://www.bfr.bund.de/de/errichtung_der_next_generation_sequenzierung_fuer_die_genomanalyse_von_bakteriellen_erregern_in_europa_engage_-202736.html

Externe Links zu Next Generation Sequencing

Weltweite Initiative *Global Microbial Identifier* zum Aufbau einer DNA Genom Datenbank zur Identifizierung und Diagnostik mikrobieller Krankheitserreger, an dem auch das BfR mitarbeitet

<https://www.globalmicrobialidentifier.org/>



„Stellungnahmen-App“ des BfR

2 Referenzen

Allard MW, Luo Y, Strain E, Pettengill J, Timme R, Wang C, Li C, Keys CE, Zheng J, Stones R, Wilson MR, Musser SM, Brown EW. 2013. On the evolutionary history, population genetics and diversity among isolates of *Salmonella* Enteritidis PFGE pattern JEGX01.0004. PLoS One 8(1), e55254.

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.

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.

Dallman T, Inns T, Jombart T, Ashton P, Loman N, Chatt C, Messelhaeusser U, Rabsch W, Simon S, Nikisins S, Bernard H, le Hello S, Jourdan da-Silva N, Kornschöber C, Mossong J, Hawkey P, de Pinna E, Grant K, Cleary P. 2016. Phylogenetic structure of European *Salmonella* Enteritidis outbreak correlates with national and international egg distribution network. *Microb. Genom* 2(8), e000070.

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 Unit States* 114(10):E1904 LP-E1912.

Deng X, Desai, PT, den Bakker, HC., Mikoleit M, Tolar B, Trees E, Hendriksen RS, Frye JG, Porwollik S, Weimer BC, Wiedmann M, Weinstock GM, Fields PI, McClelland M. 2014. Genomic epidemiology of *Salmonella enterica* serotype Enteritidis based on population structure of prevalent lineages. *Emerg Infect Dis* 20, 1481–1489.

Elson R, Awofisayo-Okuyelu A, Greener T, Swift C, Painset A, Amar C, Newton A, Aird H, Swindlehurst M, Elviss N, Foster K, Dallman TJ, Ruggles R, Grant K. 2019. Utility of WGS to describe the persistence and evolution of *L. monocytogenes* strains within crabmeat processing environments linked to outbreaks. *J Food Protect* 82:30-38.

Gillesberg Lassen S, Ethelberg S, Björkman JT, Jensen T, Sørensen G, Kvistholm Jensen A, Sørensen G, Kvistholm Jensen A, Müller L, Nielsen EM, Mølbak K. 2016. Two listeria outbreaks caused by smoked fish consumption-using whole genome sequencing for outbreak investigations. *Clin Microbiol Infect* 22 (7), 620–624.

Jagadeesan B, Gerner-Smidt P, Allard MW, Leuillet S, Winkler A, Xiao Y, Chaffron S, Van Der Vossen J, Tang S, Katase M, McClure P, Kimura B, Ching Chai L, Chapman J, Grant K. 2019. The use of next generation sequencing for improving food safety: Translation into practice. *Food Microbiol* 79:96-115.

Kovac J, den Bakker H, Carroll LM, Wiedmann M.. 2017. Precision food safety: A systems approach to food safety facilitated by genomics tools. *Trends An Chem* 96:52-61.

Lee H, Popodi E, Tang H., Foster PL. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci Unit. States* 109: E2774-E2783;

Leekitcharoenphon P, Sørensen G, Löfström C, Battisti A, Szabo I, Wasyl D, Slowey R, Zhao S, Brisabois A, Kornschöber C, Kärssin A, Szilárd J, Černý T, Svendsen CA, Pedersen K, Aarestrup FM, Hendriksen RS. 2019. Cross-Border transmission of *Salmonella* Choleraesuis var. Kunzendorf in European pigs and wild boar: infection, genetics, and evolution. *Front Microbiol.* 10:179.

Pightling, AW, Pettengill, JB., Luo Y, Baugher JD, Hugh 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

Ronholm J, Nasheri JN, Petronella N, Pagotto F. 2016. Navigating microbiological food safety in the era of whole-genome sequencing. *Clin Microbiol Rev* 29 (4):837-57.

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 (4), 350–354.

Lappi V, Wolfgang WJ, Lapierre P, Palumbo MJ, Medus C, Boxrud D. 2015. Characterization of foodborne outbreaks of *Salmonella enterica* serovar Enteritidis with Whole-Genome Sequencing Single Nucleotide Polymorphism-based analysis for surveillance and outbreak detection. *J Clin Microbiol* 53(10):3334-40.

World Health Organization (WHO). 2018. Whole genome sequencing for foodborne disease surveillance Landscape paper. ISBN 978-92-4-151368-9.

Über das BfR

Das Bundesinstitut für Risikobewertung (BfR) ist eine wissenschaftlich unabhängige Einrichtung im Geschäftsbereich des Bundesministeriums für Ernährung und Landwirtschaft (BMEL). Es berät die Bundesregierung und die Bundesländer zu Fragen der Lebensmittel-, Chemikalien- und Produktsicherheit. Das BfR betreibt eigene Forschung zu Themen, die in engem Zusammenhang mit seinen Bewertungsaufgaben stehen.