

Allgemeine Informationen

Das NRZ für Staphylokokken und Enterokokken führt seit einiger Zeit Ganzgenomsequenzierung bakteriellen Erbguts durch. Die so gewonnenen Daten sollen der hochauflösenden Diskriminierung der eingesandten Isolate dienen und zusätzliche Informationen bezüglich deren genotypischen Eigenschaften liefern, welche mittels des Standardportfolios NRZ nicht abgedeckt werden können. Hierzu wird die DNA einer reinen bakteriellen Kultur isoliert. In der hausinternen Sequenzierereinheit MF2 werden die Proben im Anschluss für die Sequenzierung vorbereitet und sequenziert. Standardmäßig verwenden wir die Illumina® Technologie im paired-end Modus mit einer Leseweite von 250 - 300 bp pro DNS-Sequenz (read). Die Auswertung und Interpretation der Rohdaten werden final am NRZ durchgeführt und in Berichten den Einsendern überstellt.

Generell erfolgt die Auswertung mit Hilfe von Analyseskripten (Kommandozeilen-basiert) und unter Zuhilfenahme der Software SeqSphere+ (Ridom GmbH, Münster). SeqSphere+ ermöglicht die Extraktion von Sequenztypen (ST), Komplextypen (CT) und von Resistenz- und Virulenzgenen, sofern eine entsprechende Datenbank hinterlegt wurde.

Für die Interpretation der Datensätze wird die Verwandtschaftsbeziehung im paarweisen Vergleich (minimum spanning tree) oder in phylogenetischen Bäumen (neighbor joining tree) dargestellt.

Es wird vorab darauf hingewiesen, dass **Ganzgenomvergleiche eine verbesserte mikrobiologische Analysemethode** darstellen, die Ergebnisse allerdings nie losgelöst von sinnvoll erhobenen und gut zusammengestellten epidemiologischen und klinischen Daten betrachtet werden sollten. Es gilt, das ganzheitliche Bild zu interpretieren. **Wir warnen davor, alleinig anhand von Genomdaten und deren Vergleichen epidemiologische Zusammenhänge zu erschließen.**

1.) Wie läuft die Ganzgenomtypisierung ab, welches Verfahren wird dabei angewandt?

Die Typisierung erfolgt auf Basis der Ganzgenomsequenz, welche mittels Illumina Sequenzier-Technologie ermittelt wurde (siehe oben). Die Genomsequenz wird mit Hilfe der, aus der Sequenzierung, generierten kurzen DNS Fragmente (reads) rekonstruiert. Diese Fragmente werden zu größeren Fragmenten zusammengesetzt (Assembly), welche im Anschluss z.B. mittels der **Software SeqSphere+** die Extraktion von **Sequenztyp (ST)** und **Komplextyp (CT)** ermöglichen. Bei der Bestimmung des ST werden nur 7 Gene betrachtet, bei der hochauflösenden CT-Bestimmung werden >1400 Gene (Spezies-spezifisch) zugrunde gelegt. Nach aktuellen Auswerteparametern sind Isolate, welche dem gleichen CT zugeordnet werden können, als nah verwandt zu betrachten und deuten auf einen Zusammenhang hin (für weiterführende Informationen siehe <https://www.ridom.de/seqsphere/>).

2.) Was ist der Unterschied zwischen einem „minimum spanning tree“ und einem „neighbor joining tree“?

Der „**minimum spanning tree**“ (MST) basiert auf einer Distanzmatrix der Isolate zueinander. In der graphischen Darstellung wird der Datensatz im paarweisen Vergleich präsentiert. Legt man, wie im Falle der Software SeqSphere+, der Berechnung x Kerngenomgene (cgMLST) zugrunde, entsprechen die Zahlen auf den Verbindungsstrichen der Isolate dem **Unterschied in Gendifferenzen** ("Allelunterschied"). Hierbei wird ein Grenzwert von x Allelen angesetzt (Spezies-spezifisch), d.h. zwei Isolate die sich in weniger als x Genen unterscheiden, können als (eng) verwandt angesehen werden. Alles innerhalb dieses Grenzwertes wird einem **Cluster** zugeschrieben und durch die Software entsprechend koloriert. Dabei ist zu beachten, dass die Zahlen entlang der Isolate nicht

einfach addiert werden dürfen. Für weiter entfernt liegende Isolate muss die Distanzmatrix betrachtet werden. (Weiterführende Informationen entnehmen Sie bitte <https://www.ridom.de/seqsphere/>.)

Für die Berechnungen des MST werden keine Modelle der Sequenzevolution herangezogen, sodass das Resultat **keinem phylogenetischen (evolutionären) Baum** entspricht.

Auch dem „**neighbor joining tree**“ (NJT) wird eine Distanzmatrix zugrunde gelegt. Bei der Berechnung der Baumtopologie kommt jedoch ein Algorithmus mit Annahme eines **Sequenzevolutionsmodell** zum Einsatz und generiert daher einen echten **phylogenetischen Baum**. Da allerdings nur ein sogenanntes Clustering-Verfahren (Ein-Schritt-Verfahren) angewendet wird und Optimalitätskriterien nicht berücksichtigt werden, entspricht auch dieser Baum nur einer Näherung an die tatsächlichen evolutionären Verhältnisse (entnommen aus: Knoope und Müller, Gene und Stammbäume, Spektrum Verlag).

3.) Ab wann spricht man von verwandten Isolaten, ab wann von der Transmission eines Stammtyps?

Die **Definitionen von Verwandtschaft und Transmission** sind nur Näherungen und nicht als absolute Bezeichnungen zu verstehen. Hier ist es von größter Bedeutung, dass die **Ergebnisse nie losgelöst von validen epidemiologischen und klinischen Daten betrachtet** werden sollten. Zudem verschwimmen die Begrifflichkeit und deren Grenzen durch die Verwendung extrem hochauflösender Analyseverfahren. Entscheidend ist jedoch die Aussage, ob eine unmittelbare Übereinstimmung zwischen epidemiologisch verdächtigen Isolaten sehr wahrscheinlich ist bzw. bestätigt oder aber grundsätzlich ausgeschlossen werden kann. Wie in Pkt. 1 und 2 erläutert, gelten zwei Isolate als miteinander verwandt die sich in weniger als x Genen voneinander unterscheiden. Dies kann daher auch als putative Transmission eines Stammtyps definiert werden.

Neben den hier verwendeten bioinformatischen Methoden gibt es auch andere Analyseverfahren, um Verwandtschaftsbeziehungen aufzulösen, z.B. der Vergleich von Einzelnukleotidunterschieden oder SNPs (Single Nucleotide Polymorphisms). Diese Analysen sind jedoch sehr aufwendig, sind nicht standardisiert und in Bezug auf die hier zu beantwortende Frage mit geringem Mehrwert, sodass sie von uns nicht routinemäßig durchgeführt werden.

4.) Was bedeutet Klonalität?

Bei einer Klonalität handelt es sich ausschließlich um sehr nah miteinander verwandte Stämme. Der Begriff eines "Klons" oder "Stamms" ist allerdings genauso wenig klar definiert wie "eng verwandt" oder "verwandt". Auch hier ist die Berücksichtigung der epidemiologischen Daten unerlässlich (siehe auch Pkt. 3).

5.) Inwieweit kann sich das Genom eines Erregers über einen definierten Zeitraum verändern?

In der Literatur finden sich vielfältige Angaben zu **Mutationsraten** bakterieller Erreger. Diese **beziehen sich ausschließlich auf das Kerngenom** und schließen Adaptionen, welche durch mobile genetische Elemente und Rekombination größerer DNA-Fragmente verursacht werden können, aus. Es wurden leicht variierende Mutationsraten für Staphylokokken und Enterokokken festgestellt. Man kann jedoch konstatieren, dass sich das **Kerngenom** der beiden Erreger um **ca. 5 SNPs pro Jahr** verändern kann. Literatur: Howden BP et al.: Genomic Insights to Control the Emergence of Vancomycin-Resistant Enterococci, mBio 4(4):e00412-13. doi:10.1128/mBio.00412-13