



**TECHNISCHE UNIVERSITÄT MÜNCHEN**

**Lehrstuhl für Mikrobielle Ökologie**

**Speziesabgrenzung innerhalb der Gattung *Pseudomonas* und  
*Bacillus cereus* s. l. auf der Basis von Genomsequenzen**

**Christopher Oliver Huptas**

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

**Doktors der Naturwissenschaften**

genehmigten Dissertation.

Vorsitzender: Prof. Dr. Rudi F. Vogel

Prüfer der Dissertation: 1. Prof. Dr. Siegfried Scherer

2. Prof. Dr. Wolfgang Liebl

Die Dissertation wurde am 17.07.2018 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 24.09.2018 angenommen.

*für meine Frau Doris und meinen Sohn André,  
die beiden wichtigsten Menschen in meinem Leben*

# Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die bei der promotionsführenden Einrichtung dem Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der TUM zur Promotionsprüfung vorgelegte Arbeit mit dem Titel:

## **Speziesabgrenzung innerhalb der Gattung *Pseudomonas* und *Bacillus cereus* s. l. auf der Basis von Genomsequenzen**

am Lehrstuhl für mikrobielle Ökologie, ZIEL-Institute for Food and Health, Department für Grundlagen der Biowissenschaften unter der Anleitung und Betreuung durch Prof. Dr. Siegfried Scherer ohne sonstige Hilfe erstellt und bei der Abfassung nur die gemäß § 6 Ab. 6 und 7 Satz 2 angebotenen Hilfsmittel benutzt habe. Ich habe keine Organisation eingeschaltet, die gegen Entgelt Betreuerinnen und Betreuer für die Anfertigung von Dissertationen sucht, oder die mir obliegenden Pflichten hinsichtlich der Prüfungsleistungen für mich ganz oder teilweise erledigt.

Ich habe die Dissertation in dieser oder ähnlicher Form in keinem anderen Prüfungsverfahren als Prüfungsleistung vorgelegt. Ich habe den angestrebten Doktorgrad noch nicht erworben und bin nicht in einem früheren Promotionsverfahren für den angestrebten Doktorgrad endgültig gescheitert.

Die öffentlich zugängliche Promotionsordnung der TUM ist mir bekannt, insbesondere habe ich die Bedeutung von § 28 (Nichtigkeit der Promotion) und § 29 (Entzug des Doktorgrades) zur Kenntnis genommen. Ich bin mir der Konsequenzen einer falschen Eidesstattlichen Erklärung bewusst.

Mit der Aufnahme meiner personenbezogenen Daten in die Alumni-Datei bei der TUM bin ich einverstanden.

Freising, den \_\_\_\_\_.2018

---

Christopher Huptas

## Danksagung

Allen voran möchte ich mich bei meinem Doktorvater Herrn Prof. Dr. Siegfried Scherer dafür bedanken mir die Möglichkeit gegeben zu haben am Lehrstuhl für mikrobiologische Ökologie zu promovieren. Herzlichen Dank für ihr durchgehend offenes Ohr und ihr Engagement vor allem während der Endzüge dieser Arbeit.

Mein besonderer Dank gilt zudem Dr. Mareike Wenning für ihre Unterstützung und die Freiheiten, die mir im gesamten Verlauf meiner Doktorandentätigkeit gewährt wurden. Vielen Dank für die Gespräche und Ratschläge sowohl in wissenschaftlicher wie auch menschlicher Hinsicht.

Darüber hinaus richte ich meinen ausdrücklichen Dank an alle ehemaligen und derzeitigen Arbeitskollegen für ihre Hilfsbereitschaft, Kollegialität und die stets freundschaftliche Atmosphäre im Institut. Mein Dank geht vor allem an Dr. Etienne Doll, Dr. Jessika Hellinckx, Dr. Genia Lücking, Dr. Lena Staib, Dr. Marlies Böhm, Mario von Neubeck, Katharina Brauer, Anna Dettling, Christopher Meier, Annemarie Siebert, Katharina Jentsch, Romy Wecko und Angela Felsl.

Außerdem möchte ich Maria Thole und Brigitte Stable für die erheiternden Gespräche und die Erledigung bürokratischer und organisatorischer Aufgaben danken.

Vergessen werden darf auch nicht Dr. Christine Wurmser, die sofern vonnöten die Betriebsbereitschaft des Covaris Ultrasonicators sichergestellt hat.

Von Herzen danke ich meiner Mutter für ihre moralische und finanzielle Unterstützung während meines Bachelor- und Master-Studiums.

Zu guter Letzt bin ich meiner Frau Doris zu außerordentlichem Dank verpflichtet. Danke für deine Liebe, deine Geduld und all die zusätzlichen Mühen, die du auf dich genommen hast, damit diese Arbeit verwirklicht werden konnte.

# Inhaltsverzeichnis

<b>I.</b>	<b>ZUSAMMENFASSUNG</b>	<b>4</b>
<b>II.</b>	<b>SUMMARY</b>	<b>6</b>
<b>III.</b>	<b>EINLEITUNG</b>	<b>8</b>
1.	<b>Prokaryotische Taxonomie</b>	<b>8</b>
1.1.	Das prokaryotische Spezies-Konzept	8
1.2.	Der polyphasische Ansatz bei Spezies Neubeschreibungen	8
1.3.	Ein neuer Goldstandard: Die <i>in silico</i> DNA-DNA Hybridisierung	10
1.4.	Paradigmenwechsel hin zu einer Genom-basierten Taxonomie?	12
2.	<b>PCR-freie Probenvorbereitung zur <i>de novo</i> Sequenzierung prokaryotischer Genome</b>	<b>14</b>
3.	<b>Die Gattung <i>Pseudomonas</i></b>	<b>15</b>
4.	<b>Die <i>Bacillus cereus sensu lato</i> Gruppe</b>	<b>17</b>
4.1.	Allgemeines, Taxonomie, Phylogenie und Bedeutung	17
4.2.	Virulenzfaktoren, ihre genetischen Determinanten und Funktionen	20
5.	<b>Zielsetzung der Arbeit</b>	<b>25</b>
<b>IV.</b>	<b>MATERIAL UND METHODEN</b>	<b>26</b>
1.	<b>Chemikalien, Lösungen und Puffer</b>	<b>26</b>
2.	<b>Medien</b>	<b>27</b>
3.	<b>Verwendete bakterielle Stämme und NCBI Daten</b>	<b>27</b>
3.1.	Verwendete bakterielle Stämme	27
3.2.	Verwendete NCBI Genomdaten	31
3.3.	Verwendete NCBI Sequenzdaten	34
4.	<b>Kultivierung und Lagerung bakterieller Stämme</b>	<b>34</b>
5.	<b>DNA Isolierung und Konzentrationsbestimmung</b>	<b>35</b>
5.1.	DNA Isolierung aus gram-positiven Bakterien	35
5.2.	DNA Isolierung aus gram-negativen Bakterien	36
5.3.	Bestimmung der DNA Konzentration	36
6.	<b><i>De novo</i> Genomsequenzierung und Assemblierung</b>	<b>37</b>
6.1.	Optimierung der PCR-freien Probenvorbereitung	37
6.2.	Genomsequenzierungen für die Optimierung der PCR-freien Probenvorbereitung	44
6.3.	Genomsequenzierungen für die Speziesabgrenzung	44
6.4.	Qualitätskontrolle der Rohdaten der Genomsequenzierungen	45
6.5.	<i>De novo</i> Genomassemblierung und Validierung der Assemblierungen	46
7.	<b>Remapping ungefilterter und gefilterter Sequenzierbibliotheken</b>	<b>47</b>
8.	<b>Evaluierung von Faktoren, die die <i>de novo</i> Assemblierung bakterieller Genome beeinflussen</b>	<b>48</b>

8.1.	Analysierte Faktoren	48
8.2.	Erstellung von Teil-Bibliotheken	48
8.3.	<i>De novo</i> Genomassemblierung der Teil-Bibliotheken	49
8.4.	Validierung der Genomassemblierungen	51
8.5.	Berechnung des relativen Assemblierungs-Wertes	52
<b>9.</b>	<b>Genomweite Vorhersage proteinkodierender Sequenzen</b>	<b>52</b>
<b>10.</b>	<b>Multilocus-Sequenz-Analyse (MLSA)</b>	<b>53</b>
10.1.	Verwendete MLSA Schemata und Außengruppen	53
10.2.	Screening nach MLSA Genen in Genomassemblierungen	54
10.3.	Vorbereitung und Alignment der verketteten MLSA Sequenzen	55
10.4.	Phylogenetische Analyse	56
<b>11.</b>	<b>Paarweise Ähnlichkeitsmaße</b>	<b>57</b>
11.1.	Berechnung und Vergleich von MLSA Sequenzähnlichkeiten	57
11.2.	Arten und Berechnung paarweiser genomischer Ähnlichkeitsmaße	57
<b>12.</b>	<b>Spezies-Grenzwert Ermittlung für die genomischen Ähnlichkeitsmaße ANIb und ANIm</b>	<b>60</b>
<b>13.</b>	<b>Ermittlung und Visualisierung von Spezies-Clustern</b>	<b>61</b>
<b>14.</b>	<b>Suche bekannter <i>B. cereus s.l.</i> Virulenzfaktoren</b>	<b>62</b>
14.1.	Vorhersage homologer Sequenzfamilien	62
14.2.	Untersuchte Virulenzfaktoren	63
14.3.	Screening-Prozedur	63
14.4.	Visualisierung von An-/Abwesenheitsprofilen	65
<b>15.</b>	<b>Einfache lineare Regressionsanalyse</b>	<b>65</b>
<b>V.</b>	<b>ERGEBNISSE</b>	<b>66</b>
<b>1.</b>	<b>Optimierung der PCR-freien Probenvorbereitung und Analyse von Faktoren, die die <i>de novo</i> Genomassemblierung beeinflussen</b>	<b>66</b>
1.1.	Optimierung des Protokolls zur PCR-freien Herstellung von Illumina Bibliotheken	66
1.2.	Einfluss der Insert-Länge und des GC-Gehaltes einer Bibliothek auf die Qualität der Sequenzierung	68
1.3.	Beeinflussende Faktoren der <i>de novo</i> Assemblierung bakterieller Genome	72
<b>2.</b>	<b>Spezies-Grenzwertableitung durch binäre logistische Regression</b>	<b>77</b>
<b>3.</b>	<b>Taxonomie und Charakterisierung bakterieller Spezies der Gattung <i>Pseudomonas</i> sowie der <i>B. cereus s.l.</i> Gruppe</b>	<b>79</b>
3.1.	Speziesabgrenzung innerhalb der Gattung <i>Pseudomonas</i>	79
3.2.	Speziesabgrenzung innerhalb der <i>B. cereus s.l.</i> Gruppe	102
3.3.	Verteilung und Eignung von <i>B. cereus s.l.</i> Virulenzfaktoren als Spezies-diskriminierende Marker	109
<b>VI.</b>	<b>DISKUSSION</b>	<b>129</b>
<b>1.</b>	<b>Optimierung der PCR-freien Probenvorbereitung und Analyse von Faktoren, die die <i>de novo</i> Genomassemblierung beeinflussen</b>	<b>129</b>
1.1.	Faktoren, die die Assemblierung beeinflussen	129

1.2.	Optimierte PCR-freie Probenvorbereitungen	133
<b>2.</b>	<b>Taxonomische Richtlinien</b>	<b>133</b>
<b>3.</b>	<b>Speziesabgrenzung in der Gattung <i>Pseudomonas</i></b>	<b>138</b>
3.1.	Speziesabgrenzung via MLSA	138
3.2.	Reklassifizierung beschriebener Arten	141
3.3.	Anzahl potentieller neuer Arten	142
<b>4.</b>	<b>Taxonomie der <i>B. cereus s.l.</i> Gruppe</b>	<b>143</b>
4.1.	Speziesabgrenzungen	143
4.2.	<i>B. cereus s.l.</i> Virulenzfaktoren als Spezies-diskriminierende Marker	146
4.3.	Das <i>B. cereus s.l.</i> Toxinovar System	151
<b>VII.</b>	<b>LITERATURVERZEICHNIS</b>	<b>156</b>
<b>VIII.</b>	<b>ABBILDUNGSVERZEICHNIS</b>	<b>178</b>
<b>IX.</b>	<b>TABELLENVERZEICHNIS</b>	<b>179</b>
<b>X.</b>	<b>ABKÜRZUNGSVERZEICHNIS</b>	<b>181</b>
<b>XI.</b>	<b>ANHANG</b>	<b>187</b>
<b>XII.</b>	<b>PUBLIKATIONEN UND KONFERENZBEITRÄGE</b>	<b>236</b>
1.	Publikationen	236
2.	Konferenzbeiträge	237

## I. ZUSAMMENFASSUNG

Genomsequenzen entwickeln sich zunehmend zu einer wichtigen Grundlage für bakterielle Speziesabgrenzungen, weil auf Genomsequenzen basierende Ähnlichkeitsmaße (z.B. ANI) zunehmend als Ersatz für die frühere DNA-DNA-Hybridisierung herangezogen werden. Daher war es zunächst wünschenswert, die Effektivität und Genauigkeit von Genomsequenzierungsprojekten zu optimieren. Mit dem Illumina TruSeq® DNA PCR-free Sample Preparation Protokoll konnten bisher nur Bibliotheken mit eher kurzen durchschnittlichen Insert-Längen produziert werden. Um die Repeat-Auflösungsfähigkeit gängiger Assemblierungstools zu verbessern, wurde die Probenvorbereitung dahingehend modifiziert und optimiert, dass eine Fertigung von Bibliotheken mit prolongierten Inserts ermöglicht wurde. Die Qualität einer Sequenzierung sinkt sowohl mit steigendem GC-Gehalt als auch durch eine Erhöhung der Insert-Länge. Aufgrund sich verstärkender negativer Effekte sollte deshalb bei der Sequenzierung von Genomen mit hohem GC-Gehalt auf moderate Insert-Längen zurückgegriffen werden. Der Vergleich von vier Assemblern zeigte, dass SPAdes bei der Rekonstruktion bakterieller Genome am effektivsten war. Mit diesem Assembler lassen sich ausgehend von Bibliotheken mit hoher Insert- und Read-Länge gute Assemblierungsergebnisse erzielen, sofern eine gewisse Sequenziertiefe (mindestens 30-fach) garantiert werden kann.

Für die allgemein verwendeten genomischen Indices ANI<sub>b</sub> und ANI<sub>m</sub> wurden bisher keine strikten Spezies-Grenzwerte definiert. Im Zuge dieser Arbeit ist es auf Grundlage experimenteller DNA-DNA-Hybridisierungsdaten von 127 Genompaaren aus 27 prokaryotischen Gattungen und mit Hilfe binärer logistischer Regression gelungen, folgende Spezies-Grenzwerte für drei ANI-Implementierungen vorzuschlagen: 95,25 % (ANI<sub>b1</sub>), 94,44 % (ANI<sub>b2</sub>) sowie 95,26 % (ANI<sub>m</sub>).

Diese Grenzwerte wurden zuerst auf Pseudomonaden angewendet. Einhundert im Rahmen dieser Arbeit *de novo* sequenzierte und assemblierte *Pseudomonas* Stämme (darunter 33 Typstämme) bilden zusammen mit 360 weiteren Stämmen, für die Sequenzdaten vom NCBI bezogen wurden, einen Datensatz, der ~ 93,6 % aller valide beschriebenen Nomenspezies (LPSN Stand April 2017) der Gattung umfasst. Nachdem die Phylogenie der Gattung mittels eines geläufigen MLSA Schemas rekonstruiert wurde, konnte die *Pseudomonas* Nomenspezies Taxonomie darauf abgebildet werden. Ein Vergleich der aktuellen Taxonomie mit den Genomospezies-

Clustern, die sich auf Basis mehrerer Spezies-abgrenzender Indices (MLSA, ANI<sub>b1</sub>, ANI<sub>b2</sub>, ANI<sub>m</sub> und gANI) sowie eindeutig definierter Spezies-Grenzwerte ergaben, deckte auf, dass insgesamt neun Nomenspezies (~ 5,3 % aller valide beschriebener Arten der Gattung) reklassifiziert werden müssten. Überdies sind mindestens 57 neue Genomospezies im Datensatz enthalten. Damit auch weiterhin möglichst verlässliche Speziesvorhersagen mit dem phylogenetischen Index getätigt werden können, bedarf es außerdem einer Anhebung des MLSA Spezies-Grenzwertes auf eine Sequenzähnlichkeit von 97,84 % (anstatt 97 %).

Als zweites Anwendungsbeispiel diente die *Bacillus cereus* s.l. Gruppe. 281 Genome, von denen acht in dieser Arbeit sequenziert und assembliert wurden, bildeten den *B. cereus* s.l. Datensatz. Die phylogenetischen Beziehungen innerhalb der *B. cereus* s.l. Gruppe wurden anhand eines MLSA Schemas, bestehend aus den vollständigen Sequenzen von sechs Housekeeping-Genen, ermittelt. Infolge der auf drei genomischen Indices (ANI<sub>b1</sub>, ANI<sub>m</sub>, gANI) basierenden Speziesvorhersagen konnte die *B. cereus* s.l. Phylogenie in 28 Genomospezies-Cluster unterteilt werden, wobei elf Cluster den zwölf beschriebenen *B. cereus* s.l. Arten (LPSN Stand November 2017) zugeordnet werden konnten. Bei der Analyse der Verteilung von mehr als einem Dutzend Virulenzfaktoren wurden zwar keine Spezies-spezifischen Präsenz/Absenz-Muster gefunden, doch möglicherweise könnte das Hämolysin A Gen auf Sequenzebene als Spezies Marker dienen. Drei neue putative Virulenzfaktoren wurden durch bioinformatische Analyse gefunden. Zu diesen gehören ein Hämolysin mit einer Sequenzähnlichkeit von ~ 42 % zum Hämolysin III sowie zwei Metalloproteasen aus der Proteinfamilie der Immunitätsinhibitoren.

Viele *B. cereus* s.l. Isolate werden noch heute nach ihren human- oder insektenpathogenen Eigenschaften benannt. Dies erklärt, warum weniger als ein Drittel der Stämme des untersuchten Datensatzes den phylogenetisch korrekten Artnamen trägt. Um Fehlbenennungen dieses Ausmaßes in Zukunft zu vermeiden, vereint das in dieser Arbeit vorgeschlagene Toxinovar System die phylogenetische Stellung mit dem Pathotyp eines Isolates in einer sinnvollen Nomenklatur für die gesamte *B. cereus* s.l. Gruppe.

## II. SUMMARY

Regarding bacterial species delineation, genome sequences are becoming more and more important, since DNA-DNA hybridization is increasingly being replaced by whole-genome sequence based similarity measurements (e.g. ANI). Therefore, it seems advisable to enhance the efficiency and accuracy of genome sequencing projects. Until now the Illumina TruSeq® DNA PCR-free Library Preparation Procedure is restricted to the generation of libraries with rather short average insert sizes. For better exploitation of the repeat bridging capabilities of current state of the art assemblers, sample preparation was optimized to enable the creation of sequencing libraries with prolonged inserts. However, sequencing quality of a library decreases with increased genomic CG-content or enlarged average insert size. To exclude mutually reinforcing negative effects on sequencing quality, GC-rich genomes should be sequenced only at moderate insert lengths. A comparison of four assembly tools showed that SPAdes was the assembler performing best. Using SPAdes, improved genome assemblies were obtained at high insert and read lengths as long as sequencing depths of at least 30-fold could be guaranteed.

For the commonly used genomic indices ANI<sub>b</sub> and ANI<sub>m</sub>, no strict species thresholds were defined so far. Using binary logistic regression based on DNA-DNA hybridization data of 127 genome pairs from 27 different prokaryotic genera, it was possible to propose such cutoff values for three different ANI implementations. These correspond to genome sequence similarities of 95.25 % (ANI<sub>b1</sub>), 94.44 % (ANI<sub>b2</sub>) and 95.26 % (ANI<sub>m</sub>).

Initially, the inferred species boundaries were applied to the genus *Pseudomonas*. In the course of this study one hundred *Pseudomonas* strains (including 33 type strains) were sequenced and assembled *de novo*. Together with 360 additional strains taken from the NCBI database the complete data set included ~ 93.6 % of all described and validly named species of the genus (LPSN as of April 2017). After reconstruction of the strains' evolutionary relationships using a common MLSA scheme, the current taxonomy was mapped to the phylogeny of the genus. Comparing taxonomy with genomospecies clusters based on several species delineating indices (MLSA, ANI<sub>b1</sub>, ANI<sub>b2</sub>, ANI<sub>m</sub>, gANI) and strictly defined species thresholds revealed the existence of nine species (~ 5.3 % of all *Pseudomonas* nomenclatures) which need to be reclassified. Furthermore, the data set contains at least 57 new genomospecies. To

ensure species membership predictions via the phylogenetic index to be as reliable as possible, the MLSA species cutoff value should be increased to a sequence similarity of 97.84 % (instead of 97 %).

The *Bacillus cereus s.l.* data set comprised 281 genomes of which eight were sequenced and assembled *de novo*. Phylogenetic inference was performed by means of a MLSA scheme that consists of the complete coding sequences of six housekeeping genes. Species predictions based on three genomic indices (ANi<sub>b1</sub>, ANIm, gANI) subdivided the phylogeny of the group into 28 genomospecies clusters, whereby eleven clusters corresponded to twelve described nomenclatures (LPSN as of November 2017). Although the distributions of more than a dozen virulence factors uncovered no species-specific presence/absence patterns, the *hlyA* gene might serve as a species marker at sequence level. In addition, three previously unknown virulence factors were found by bioinformatic analyses. These include a putative hemolysin that shares ~ 42 % sequence homology to hemolysin III as well as two putative immune inhibitory metalloproteases.

Still, many *B. cereus s.l.* strains are named in accordance to their human- or entomopathogenic properties. This explains why only less than one third of the analyzed isolates are carrying species names consistent to their phylogenetic affiliation. To avoid future misnomers, the toxinovar nomenclatural system which considers a strain's phylogenetic position and genomospecies membership during the assignment of names is proposed.

### III. EINLEITUNG

#### 1. Prokaryotische Taxonomie

##### 1.1. Das prokaryotische Spezies-Konzept

Das prokaryotische Spezies-Konzept ist Gegenstand ständiger Revisionen und Verfeinerungen, die in der Regel mit der Verfügbarkeit technischer Neuerungen einhergehen [1]. Das unter Taxonomen wohl am weitesten verbreitete und akzeptierte Konzept besagt: Eine prokaryotische Spezies ist eine „(preferably) genomically coherent group of individual isolates/strains sharing a high degree of similarity in (many) independent features, comparatively tested under highly standardized conditions“ (Stackebrandt *et al.* 2002) [2]. Das Konzept basiert in wesentlichen Teilen auf dem phylo-phänetischen Spezies-Konzept von Rosselló-Mora und Amann aus dem Jahr 2001 [1], das von diesen beiden Autoren im Jahr 2015 durch eine Neuformulierung ersetzt wurde: „A species is a category that circumscribes monophyletic, and genomically and phenotypically coherent populations of individuals that can be clearly discriminated from other such entities by means of standardized parameters“ (Rosselló-Mora & Amann 2015) [3]. Der Nachweis einer Gruppe von Isolaten als Monophylum mit sowohl genomischer wie auch phänotypischer Kohärenz wird durch den polyphasischen Ansatz bei Spezies Neubeschreibungen erbracht [3].

##### 1.2. Der polyphasische Ansatz bei Spezies Neubeschreibungen

Ziel einer Spezies Neubeschreibung ist die Charakterisierung einer bisher unbekannten Art sowie deren Abgrenzung zu den am nächsten verwandten, bereits valide beschriebenen Arten. Erreicht wird dieses Ziel durch einen polyphasischen Ansatz bei dem genotypische (genomische), chemotaxonomische und phänotypische Charakteristika für die Klassifizierung einer neuen Art herangezogen werden [2, 4, 5].

Mit der Einführung der DNA-DNA Hybridisierungstechnologien [6] in den 1960er Jahren wurde die genomische Kohärenz einer prokaryotischen Art daran gemessen, dass ihre Mitglieder einen DNA-DNA Reassoziierungsgrad  $\geq 70$  % aufwiesen [7-9]. In den 1970er Jahren hielt die 16S rRNA Analyse Einzug in die prokaryotische Taxonomie [10, 11]. Noch heute sind 16S rRNA Gensequenzen ein obligatorischer

Bestandteil jedweder Neubeschreibung [5, 12], was nicht zuletzt auch auf den Einzug der Sanger-Sequenzierung [13] und der Polymerase-Kettenreaktion [14] in die molekularbiologischen Labore zurückzuführen ist. Seine zentrale Bedeutung für die phylogenetische Analyse erlangte das 16S rRNA Gen, weil es ubiquitär vorhanden, hoch konserviert (sowohl in Hinsicht auf seine sekundäre Struktur wie auch molekulare Funktion) und nur selten Gegenstand horizontalen Gentransfers ist [15]. Zudem konnten Stackebrandt und Goebel zeigen, dass bei einer 16S rRNA Sequenzähnlichkeit  $< 97\%$  ein DNA-DNA Hybridisierungswert  $\geq 70\%$  zwischen zwei prokaryotischen Organismen nicht mehr zu erwarten ist [16]. Diese Erkenntnis erleichterte die taxonomische Arbeit, da arbeitsaufwendige und fehleranfällige DNA-DNA Hybridisierungsversuche im Rahmen der Speziesabgrenzung nur noch bei Sequenzähnlichkeiten  $\geq 97\%$  durchgeführt werden mussten [5]. Im späteren Verlauf wurde dieser Grenzwert unter Berücksichtigung einer belastbareren Datengrundlage auf 98,7-99 % nach oben korrigiert [17]. Zwar ist die 16S rRNA Analyse für die adäquate Einordnung einer Spezies in die höheren taxonomischen Ränge (Domäne bis hin zur Gattung) geeignet, jedoch reicht die Auflösungskraft von 16S rRNA Sequenzen auf Spezies- und Subspeziesebene in der Regel nicht mehr aus [16, 18].

Um verlässliche Aussagen über die zwischen- und innerartlichen Beziehungen von Mikroorganismen aufstellen zu können, wird häufig auf die phylogenetische Rekonstruktion mittels einzelner proteinkodierender Gene oder der Multilocus-Sequenz-Analyse (MLSA) zurückgegriffen [19-21]. Einige Zeit lang galt MLSA sogar als möglicher Kandidat die DNA-DNA Hybridisierung (DDH) zu ersetzen [19, 21, 22], jedoch konnte weder eine allgemein anwendbare Menge an Genen noch ein universell gültiger Grenzwert für die prokaryotische Speziesabgrenzung bestimmt werden [3, 19]. Vielmehr wurden unterschiedliche MLSA Schemata (mit jeweils eigenen Spezies-Grenzwerten) für bestimmte Gattungen und Gruppen von Mikroorganismen entwickelt [23-26]. Nichtsdestotrotz erweitert die MLSA das Repertoire der genotypischen Untersuchungsmöglichkeiten und dient neben der 16S rRNA Analyse (und anderen) als eine Methode zum Nachweis des monophyletischen Status und der genomischen Kohärenz einer neuen Art [3, 27].

Ursprünglich basierte die taxonomische Charakterisierung prokaryotischer Arten einzig auf der Untersuchung phänotypischer Merkmale, wie der Zell- und Koloniemorphologie als auch den allgemeinen Wachstumsbedingungen [1, 18]. Im späteren Verlauf kam die Bestimmung der physiologischen und biochemischen

Eigenschaften einer Spezies hinzu [28]. Überdies nimmt die Chemotaxonomie bei heutigen Neubeschreibungen eine zentrale Rolle ein. Hierbei handelt es sich um die chemische Charakterisierung der äußeren Zellschicht (Peptidoglykan, Teichonsäuren, Mykolsäuren, etc.), der Zellmembran (Fettsäuren, polare Lipide, respiratorische Lipochinone, Pigmente, etc.) und des Zytoplasmas (Polyamine) [5]. Die fortschreitenden Entwicklungen im Bereich der massenspektrometrischen Techniken ermöglichen den Nachweis der phänotypischen Kohärenz einer Art auch im Hochdurchsatzverfahren [3, 29, 30]. Als vielversprechend gelten dabei die Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry (MALDI-TOF MS) [31] sowie die Ion Cyclotron Resonance-Fourier Transform Mass Spectrometry (ICR-FT MS). Der Vorteil solcher massenspektrometrischer Verfahren liegt im Aufbau metabolischer Datenbanken, die für taxonomische Zwecke abgeglichen werden können [27]. In diesem Zusammenhang ist auch die Fourier Transform Infrared Spectroscopy (FTIR Spectroscopy) als eine leistungsstarke Methode zur schnellen Identifizierung und Diskriminierung mikrobieller Spezies anzusehen [32].

### **1.3. Ein neuer Goldstandard: Die *in silico* DNA-DNA Hybridisierung**

Ungefähr 50 Jahre lang galt die DDH als der Goldstandard für die Abgrenzung prokaryotischer Arten [5, 9]. Der entscheidende Nutzen der DDH lag darin, dass sich prokaryotische Genome in klar voneinander getrennte Cluster (= Spezies) separieren ließen [33, 34]. Allerdings gelten DDH Experimente heute aufgrund ihrer arbeitsaufwendigen und fehleranfälligen Natur als weitgehend überholt.

Im Jahr 1995 wurde das erste bakterielle Genom (*Haemophilus influenzae* Rd) sequenziert und assembliert [35]. Zu dieser Zeit waren Genomsequenzierungen noch kostspielig und zeitaufwendig, da sie auf der Sanger-Sequenzierung und der bakteriellen Klonierung von DNA-Fragmenten beruhten [36, 37]. In der zweiten Hälfte der 2000er Jahre kamen die Next-Generation-Sequencing (NGS) [38-40] und Third-Generation-Sequencing (TGS) [41, 42] Technologien auf. Diese ermöglichten die schnelle und kostengünstige Sequenzierung ganzer Genome [37, 43]. Stetige Neu- und Weiterentwicklungen der NGS/TGS Sequenzierplattformen [44, 45], Verbesserungen bei den Probenvorbereitungsprotokollen [37] und die Etablierung eines breiten Angebotes kommerzieller Sequenzierdienstleistungen führten dazu, dass heute selbst allgemeine (nicht-spezialisierte) mikrobiologische Labore diese Technologien weitläufig nutzen können [36]. Die Verfügbarkeit einer kontinuierlich

steigenden Anzahl an prokaryotischen Genomsequenzen, förderte die Suche nach einem *in silico* Pendant zur DDH [3, 36].

Die Average Nucleotide Identity (ANI) war der erste Index der als Ersatz für die DDH vorgeschlagen wurde [46]. Ursprünglich basierte der Algorithmus zur Berechnung der durchschnittlichen Ähnlichkeit zwischen zwei prokaryotischen Genomen auf der Menge ihrer gemeinsamen (orthologen) Gene [46]. Später wurde der Algorithmus dahingehend modifiziert, dass alle miteinander geteilten Genomregionen zur Berechnung der Ähnlichkeit herangezogen wurden [47, 48]. Welche der genomischen Sequenzbereiche als zueinander homolog angesehen werden (und somit in die Berechnung der Ähnlichkeit mit eingehen) wurde mittels Standard Nucleotide BLAST [49] (ANIb) oder der MUMmer Software [50] (ANIm) bestimmt. Korrelationsanalysen ergaben, dass ein ANIb/ANIm Wert im Bereich von 95-96 % einem DNA-Reassoziierungsgrad  $\geq 70$  % entspricht [47, 48]. Unter Taxonomen sind JSpecies [48] und JSpeciesWS (<http://jspecies.ribohost.com/jspeciesws/>) [51] die wohl am häufigsten genutzten Tools zur Berechnung des ANIb und ANIm. Seit der Einführung der beiden Indices als neue *in silico* Goldstandards zur prokaryotischen Speziesabgrenzung [48], wurden noch weitere ANI-Varianten (gANI [52], OrthoANI [53], OrthoANLu [54], FastANI [55], etc.) entwickelt. Dabei unterscheiden sich die einzelnen Algorithmen in der Art der Sequenzen (Gensequenzen vs. Genomsequenzen) auf denen die Berechnung des evolutionären Verwandtschaftsgrades zweier Genome beruht, sowie dem Typus der Sequenzanalyse (Alignment-basiert vs. Alignment-frei). Je nach algorithmischer Ausrichtung können sich teils drastische Unterschiede in der benötigten Rechenzeit bei großen Datensätzen ergeben [54, 55].

Ein weiterer häufig verwendeter Index zum Zwecke der Speziesabgrenzung ist die Genome-Blast-Distance-Phylogeny (GBDP) [56-59]. Während es sich bei den verschiedenen ANI-Varianten um zwischengenomische Ähnlichkeitsmaße handelt, ist die GBDP ein Maß für die Distanz zweier Genome. Die Genome-to-Genome Distance Calculator (GGDC) Software (<http://ggdc.dsmz.de/>) errechnet den Distanzwert eines Genompaars und transformiert ihn in einen digitalen DNA-DNA Hybridisierungswert (dDDH). Wie im Falle der (experimentellen) DDH liegt der Spezies-Grenzwert der dDDH ebenfalls bei 70 %.

#### 1.4. Paradigmenwechsel hin zu einer Genom-basierten Taxonomie?

Einige der Grundvoraussetzungen, der polyphasische Ansatz selbst wie auch die aktuell vollzogene Praxis bei der Beschreibung neuer prokaryotischer Taxa sind seit geraumer Zeit Gegenstand von Kritik [18, 60-62]. So fordert die gängige Praxis, dass jede neue Art über diskriminative phänotypische (inkl. chemotaxonomische) Merkmale verfügt [5]. Die experimentelle Charakterisierung des Phänotyps ist jedoch zeitaufwendig, da entsprechende Tests in der Regel mit den Typstämmen der am nächsten verwandten Arten (Referenzstämmen) zeitgleich ausgeführt werden müssen [60, 62]. Außerdem unterscheiden sich viele der Tests im Grad ihrer Reproduzierbarkeit und Vergleichbarkeit [18, 27] und müssen häufig unter hoch standardisierten Bedingungen durchgeführt werden [18]. Hinzu kommt, dass die meisten Neubeschreibungen auf Grundlage einzelner oder weniger Stämme erfolgen [29, 63, 64]. Derartige Einzelstamm-Beschreibungen können die intraspezifische Variabilität des Phänotyps jedoch nicht akkurat abbilden [3], womit die diagnostische Aussagekraft der ermittelten Spezies-spezifischen phänotypischen Charakteristika von vornherein in Frage zu stellen ist [29, 60].

Die polyphasische Taxonomie unter der Prämisse einer möglichst umfangreichen Charakterisierung neuer Taxa [5] kann nicht mehr mithalten mit der rapide steigenden Anzahl prokaryotischer Genomsequenzen in den öffentlichen Datenbanken [36, 62, 65]. Es ist anzunehmen, dass eine zunehmende Diskrepanz zwischen der in den Datenbanken befindlichen Spezies-Diversität und der Anzahl bisher valide beschriebener Arten aufkommen wird. Ende 2016 enthielt die Genomes OnLine Database (GOLD) [66] ungefähr 67.500 prokaryotische Genomsequenzierungseinträge mit dem Status 'Complete and Published' oder 'Permanent Draft' (<https://gold.jgi.doe.gov/>). Fünf Jahre zuvor waren es dahingegen lediglich um die 7.400. Im gleichen Zeitraum stieg der Umfang publizierter Artnamen mit Aufnahme in die List of Prokaryotic names with Standing in Nomenclature (LPSN) [67, 68] jedoch nur von 11.070 auf 15.267 (<http://www.bacterio.net/-number.html>). Zu den drei traditionellen Hauptaufgabengebieten der prokaryotischen Taxonomie zählen die Klassifizierung, die Benennung und die Identifizierung von Mikroorganismen [69]. Daneben wurde ein weiteres viertes Ziel bereits vorgeschlagen, nämlich die Erfassung der gesamten prokaryotischen Diversität [70]. Schätzungen gehen von über einer Million prokaryotischer Spezies in der Biosphäre aus [71, 72]. Unter Beibehaltung der momentanen Klassifikationsgeschwindigkeit (~

1.000 valide beschriebene Arten pro Jahr [68]) würden Jahrhunderte bis Jahrtausende benötigt, um die weltweite Diversität abzudecken [27, 62]. Ein weiteres Hindernis ergibt sich in diesem Zusammenhang mit der Regel 18a des International Code of Nomenclature of Bacteria, nach der der Typstamm einer neuen Spezies als lebende Reinkultur vorliegen soll [73]. Aus heutiger Sicht erscheint eine Beschränkung der Taxonomie auf den kultivierbaren Anteil prokaryotischer Arten aus zweierlei Gründen antiquiert. Erstens, die meisten Prokaryoten können noch nicht oder nur schwer kultiviert werden [74, 75]. Zweitens, kultivierungsunabhängige Methoden wie das Single Cell Genome Sequencing ermöglichen mittlerweile die Rekonstruktion prokaryotischer Genome ausgehend von nur einer einzigen Zelle [76, 77].

Vor kurzem wurde vorgeschlagen, dass alternativ zur Reinkultur auch die Genomsequenz (inkl. genomischer DNA) eines Typstammes als Typmaterial zur Aufbewahrung in den internationalen Stammsammlungen hinterlegt werden könnte [78]. Eine entsprechende Änderung der Regel 18a wäre ein erster Schritt in Richtung hin zu einer Genom-basierten Taxonomie. Tatsächlich lassen sich viele der zentralen Bestandteile der polyphasischen Taxonomie, wie z.B. die 16S rRNA Analyse, die DDH (in Form einer *in silico* DHH) und die Bestimmung des GC-Gehaltes allein auf Grundlage von Genomsequenzen realisieren [29, 36], solange diese über einen ausreichenden Qualitätsstandard verfügen [12]. Zudem sind dank der Anstrengungen von Genomsequenzierungsinitiativen, wie der Genomic Encyclopedia of Bacteria and Archaea (GEBA) [79], immer mehr Genomdaten von Typstämmen in den öffentlichen Datenbanken anzufinden [80], auf die im Rahmen taxonomischer Arbeiten zurückgegriffen werden kann. Darüber hinaus wird die Genomsequenzierung des Typstammes im Rahmen von Neubeschreibungen von vielen Journalen bereits dringend empfohlen oder sogar gefordert [3]. In absehbarer Zeit sollten somit die Genomsequenzen der meisten prokaryotischen Typstämme zur Verfügung stehen.

Taxonomen unterschiedlicher Couleur befinden sich im Diskurs darüber, welchen Stellenwert die phänotypische Charakterisierung in einer Genom-basierten Taxonomie noch einnehmen soll? Hier reichen die Ansätze von einer *in silico* Vorhersage des Phänotyps [62, 81, 82] über seine minimale experimentelle Bestimmung [60, 61] bis hin zum Aufbau von interaktiven Metabolom-Datenbanken [3, 27]. Ungeachtet des Ausgangs dieses Diskurses waren es auch dieses Mal

technische Neuerungen (NGS/TGS) die den Paradigmenwechsel einleiteten. Was einst formuliert wurde scheint (bald) Realität zu sein, nämlich “that the complete deoxyribonucleic acid (DNA) sequence would be the reference standard to determine phylogeny and that phylogeny should determine taxonomy. Furthermore, nomenclature should agree with (and reflect) genomic information“ (Wayne *et al.* 1987) [9].

## **2. PCR-freie Probenvorbereitung zur *de novo* Sequenzierung prokaryotischer Genome**

Die Next-Generation-Sequencing (NGS) Technologien revolutionierten das Feld der molekulargenetischen Forschung [83]. Heute gehört die Hochdurchsatzsequenzierung der Firma Illumina zu den vorherrschenden NGS Technologien [84].

In der Vergangenheit wurden mehrfach Anstrengungen unternommen bereits etablierte Probenvorbereitungsprotokolle zur Herstellung von paired-end Bibliotheken für die Genomsequenzierung zu verbessern [85-88]. Momentan stellt das kommerziell erhältliche Illumina TruSeq® DNA PCR-free Sample Preparation Protokoll die wahrscheinlich am meisten genutzte Probenvorbereitungsprozedur zu diesem Zwecke dar. Es macht Gebrauch von der Covaris Adaptive Focused Acoustics™ Technologie, die im Gegensatz zum enzymatischen Verdau eine randomisierte Fragmentierung der genomischen DNA garantieren kann. Diese Art der ungerichteten DNA Scherung sowie die Einsparung der PCR zur Anreicherung doppelt ligierter DNA Fragmente (Enrichment-PCR) führt zu einer signifikanten Verringerung ungleicher genomischer Sequenziertiefen [86, 89]. Hinzu kommt, dass aufgrund der Verwendung magnetischer Beads im Zuge der DNA Aufreinigung und Längenselektion die Kontaminationsgefahr geringer als bei den traditionellen Gel-basierten Methoden ausfällt. Einer der Nachteile PCR-freier Probenvorbereitungen liegt in der benötigten Menge an DNA Ausgangsmaterial (1-4 µg), womit die Anwendung PCR-freier Protokolle auf bakterielle Stämme, die unter Laborbedingungen gut zu kultivieren sind, eingeschränkt ist.

Mit dem kommerziellen Probenvorbereitungsprotokoll ist die Herstellung von nicht mehr als zwei Sequenzierbibliotheken mit eher kurzen Fragmentlängen von 350 bp bzw. 550 bp im Durchschnitt möglich. Allerdings ist dies im Falle eukaryotischer Genome ungünstig, da die entsprechenden Sequenzierungsprojekte zur Steigerung

der Qualität einer Assemblierung Gebrauch von mehreren Bibliotheken mit variierenden Insert-Längen machen könnten [90, 91]. Andererseits war eines der wesentlichen Erkenntnisse des GAGE-B (Genome Assembly Gold-standard Evaluation for Bacteria) Wettbewerbs, dass bemerkenswert gute Assemblierungen der vergleichsweise viel kleineren prokaryotischen Genome auf Basis einer einzelnen Illumina Sequenzierbibliothek möglich sind, solange diese über eine ausreichend hohe Sequenziertiefe verfügt [92]. Um die Leistungsfähigkeit derartiger Assemblierungen noch weiter anzuheben wird eine Prolongation der durchschnittlichen Insert-Länge prokaryotischer Sequenzierbibliotheken für sinnvoll erachtet. Dies könnte die Fähigkeit der gängigen Assemblierungstools zur Überbrückung und Auflösung von Repeatstrukturen verstärken [93]. Die Generierung von Genomsequenzen ausgehend von nur einer einzigen Illumina Bibliothek schränkt den Parameterraum, in dem eine Leistungssteigerung der Assemblierung erreicht werden kann, stark ein. In diesem Zusammenhang zählen zu den wichtigsten Faktoren: die Qualität der Read-Daten [94], die Read-Länge [90], die erreichte Sequenziertiefe [95-97] und die verwendete Assemblierungssoftware (inkl. Parameter Tuning) [92, 94, 98, 99]. In Bezugnahme zur Beschaffenheit der Insert-Längenverteilung einer Sequenzierbibliothek ist allerdings nur wenig bekannt.

### **3. Die Gattung *Pseudomonas***

Die Gattung *Pseudomonas* wurde vom deutschen Botaniker Walter Migula im Jahr 1894 beschrieben [100, 101] und beinhaltet gerade oder leicht gekrümmte, gram-negative, aerobe, nichtsporulierende Stäbchenbakterien, die in der Regel ein bis mehrere polare Flagellen, keine Polyhydroxybutyrat-Granula und einen GC-Gehalt von 58-69 % aufweisen [102]. In ihrer mehr als hundertjährigen Geschichte unterlag die Gattung einigen taxonomischen Überarbeitungen [103-105]. Eine der bedeutenderen Revisionen erfolgte Mitte der 1980er Jahre auf Grundlage von rRNA-DNA Hybridisierungsversuchen, wonach sich die Gattung in die rRNA-Homologie Gruppen I bis V aufspaltete [106, 107]. Weiterführende rRNA-DNA Hybridisierungsstudien untermauerten zwar die Existenz der fünf Gruppen, jedoch wurde festgestellt, dass diese genotypisch zu weit auseinander liegen um eine einzige Gattung repräsentieren zu können [108-110]. Aufgrund der Ansässigkeit der *Pseudomonas* Typspezies (*P. aeruginosa*) [111, 112] wurde eine Eingrenzung der Gattung auf die in der rRNA Gruppe I befindlichen Spezies vorgeschlagen [108-110], was heute allgemein anerkannt wird [103, 113]. Ergänzend zur rRNA-DNA

Hybridisierung erfolgte mittels 16S rRNA Katalogisierung und der Analyse repräsentativer 16S rRNA Sequenzen die Einordnung von *Pseudomonas sensu stricto* (= rRNA Gruppe I) in Woesses  $\gamma$ -Subdivision der *Proteobacteria* [110, 114-116] (heute: *Gammaproteobacteria* [117]). Die auf der rRNA-DNA Hybridisierung beruhende Gliederung von *Pseudomonas* s.s. in drei Untergruppen [107] wurde im späteren Verlauf durch die phylogenetische Analyse mit (nahezu) kompletten 16S rRNA Sequenzen im Grundsatz bestätigt [118, 119]. Überdies wurde festgestellt, dass sich die Gattung in zwei große intragenerische Cluster, nämlich die *P. aeruginosa* sowie die *P. fluorescens* Abstammungslinie, aufspaltet [118]. Diese Beschaffenheit der *Pseudomonas* Phylogenie konnte zudem auf Basis von Fettsäuremethylester-Profilen sowie der Analyse zweier konkatenierter Housekeeping-Gene verifiziert werden [120, 121]. Von allen drei Ansätzen wies die phylogenetische Rekonstruktion unter Verwendung der beiden partiellen Housekeeping-Gene *gyrB* (DNA gyrase subunit B) und *rpoD* (RNA polymerase subunit D) die höchste zwischenartliche Auflösungskraft auf [118, 120, 121]. Auch andere proteinkodierende Gene wurden im Zuge phylogenetischer und taxonomischer Studien der Gattung genutzt [122-124]. So konnte z.B. gezeigt werden, dass im Vergleich zur 16S rRNA das *rpoB* (RNA polymerase subunit B) Gen eine fast dreifach höhere Auflösungskraft besitzt [123]. Im Jahr 2010 publizierten Mulet *et al.* ein MLSA Schema bestehend aus den partiellen Sequenzen der vier Gene 16S rRNA, *gyrB*, *rpoD* sowie *rpoB* und schlugen einen MLSA Spezies-Grenzwert bei einer Sequenzähnlichkeit von 97 % vor. Des Weiteren konnte festgestellt werden, dass die Konkatenation der ersteren drei Gene (16S rRNA, *gyrB* und *rpoD*) für eine verlässliche phylogenetische Analyse der Gattung ausreichend ist und dass die zusätzliche Verwendung des *rpoB* Genes bei der Diskriminierung von Typstämmen keine Erhöhung der Auflösungskraft mit sich bringt [25]. Im Rahmen einer darauf folgenden Erweiterung und Aktualisierung der *Pseudomonas* Phylogenie wurde der MLSA Spezies-Grenzwert nochmals durch einen Abgleich mit den Spezies-diskriminierenden Indices ANIb, ANIm und GBDP bestätigt [125]. Außerdem bezeugten die MLSA Phylogenien die bereits beobachtete Aufteilung der Gattung in zwei große intragenerische Cluster [25, 118, 120-122, 125]. Später konnte neben der *P. aeruginosa* und *P. fluorescens* Abstammungslinie noch ein drittes Cluster (die *P. pertucinogena* Linie) definiert werden [105, 126]. Zusätzlich zu den Sequenz-basierten Methoden (z. B. MLSA) [127-130] und genomischen Ansätzen (z.B. ANI, GBDP) [125, 131, 132] kamen für die Charakterisierung, Identifizierung und

Differenzierung von *Pseudomonas* Spezies auch phänotypische Techniken zum Einsatz. Dazu gehörten unter anderem die Fluoreszenzspektroskopie [133], die Siderotypisierung [134-136] und MALDI-TOF MS [128, 130, 137].

Die Vertreter der Gattung *Pseudomonas* zeichnen sich durch ihr ubiquitäres und weltweites Vorkommen aus [104, 105]. Derzeit (April 2017) sind 171 valide beschriebene *Pseudomonas* Arten in der LPSN verzeichnet [67, 68]. Diese wurden hauptsächlich aus Böden (z.B. *P. bauzanensis*), Boden-assoziierten Habitaten, wie der Rhizosphäre (z.B. *P. rhizosphaerae*) oder Phyllosphäre (z.B. *P. graminis*), und Gewässern (z.B. *P. litoralis*) isoliert [104, 105, 138-141]. Außerdem zählen extreme Lebensräume, wie heiße Quellen (z.B. *P. guguanensis*), Wüsten (z.B. *P. arsenicoxydans*), die Tiefsee (z.B. *P. oceani*) oder die Antarktis (z.B. *P. antarctica*), zu den Isolationsorten einiger Spezies [142-145]. Darüber hinaus beinhaltet die Gattung human- (z.B. *P. aeruginosa*), tier- (z.B. *P. anguilliseptica*), pflanzen- (z.B. *P. syringae*) und pilzpathogene (z.B. *P. tolaasii*) Arten [111, 146-150]. Die immense metabolische und physiologische Vielfalt der Gattung [151, 152] bildet außerdem die Grundlage für die biotechnologische Nutzung von *Pseudomonas* spp. im Zuge der Förderung des Pflanzenschutzes und -wachstums (z.B. *P. fluorescens*), der biologischen Sanierung der Umwelt (z.B. *P. aeruginosa*), der Herstellung von Bioplastik (z.B. *P. putida*) und der Produktion von Biotensiden (z.B. *P. aeruginosa*), um nur einige der möglichen Anwendungsgebiete zu nennen [153-158].

## 4. Die *Bacillus cereus sensu lato* Gruppe

### 4.1. Allgemeines, Taxonomie, Phylogenie und Bedeutung

Bei der *Bacillus cereus sensu lato* (s.l.) Gruppe handelt es sich um einen Spezies Komplex von gram-positiven, fakultativ anaeroben, endosporenbildenden Stäbchenbakterien mit geringem GC-Gehalt. *B. cereus* s.l. ist ein weit verbreitetes bodenbewohnendes Bakterium mit saprophytischem Lebenszyklus [159, 160]. Der Komplex umfasst insgesamt neun Arten, die hauptsächlich auf Grundlage phänotypischer und pathogener Charakteristika etabliert wurden. Neben der namensgebenden Art *B. cereus sensu stricto* (s.s.) [161], gehören auch *B. anthracis* [162], *B. thuringiensis* [163], *B. mycoides* [164], *B. pseudomycoides* [165], *B. weihenstephanensis* [166], *B. cytotoxicus* [167], *B. toyonensis* [168] und *B. wiedmannii* [169] zur Gruppe. Zudem werden *B. "gaemokensis"* [170], *B. "manliponensis"* [171] und *B. "bingmayongensis"* [172] als Mitglieder des Komplexes

angesehen, wobei die Namen dieser drei Arten (noch) nicht in die LPSN aufgenommen wurden. Die Mitglieder des Spezies Komplexes weisen eine große Variationsbreite hinsichtlich ihres pathogenen Potentials und ihrer agrarökonomischen Bedeutung auf [160, 173, 174].

*B. anthracis* ist das ätiologische Agens der schwerwiegenden Milzbranderkrankung bei Säugetieren [175, 176] und wird als einer der potentesten biologischen Kampfstoffe angesehen [177]. *B. thuringiensis* (Bt) produziert eine große Bandbreite an  $\delta$ -Endotoxinen (Cry und Cyt) mit Wirkung gegen ein weites Spektrum von Schadinsekten [178]. Bt-Produkte gehören zu den erfolgreichsten und umsatzstärksten Biopestiziden weltweit [174, 179]. *B. cereus* s.s. ist verantwortlich für zwei Arten von Lebensmittelerkrankungen, die entweder durch das Auftreten von Emesis oder Diarrhoe gekennzeichnet sind. Beim emetischen Syndrom handelt es sich um eine Intoxikation mit dem Gift Cereulid, das im kontaminierten Lebensmittel gebildet wird. Die hervorgerufenen Hauptsymptome sind Übelkeit und Erbrechen [180-183]. Dahingegen ist die Diarrhoe Verlaufsform zurückzuführen auf eine Toxikoinfektion. Durch das Wirken von ein oder mehreren Enterotoxinen im Dünndarm kommt es zu Bauchschmerzen und Durchfall. Normalerweise verlaufen beide gastrointestinale Erkrankungsformen in einem Zeitraum von höchstens 24 Stunden selbstbegrenzend [181, 183-185]. Jedoch konnten auch schon sehr viel schwerwiegendere Lebensmittelerkrankungen auf emetische *B. cereus* s.s. zurückgeführt werden [186-190]. Zudem sind Erkrankungsfälle bekannt, die durch die (milden) Symptome beider Verlaufsformen gekennzeichnet waren [180]. Neben seiner stetig steigenden Bedeutung als Lebensmittelpathogen [191-193], ist *B. cereus* s.s. auch ein ernst zu nehmender opportunistischer Infektionserreger [194-200]. Die beiden Arten *B. mycoides* und *B. pseudomycoides* zeichnen sich durch ihr rhizoidales Koloniewachstum aus [164, 165, 201]. Neben *B. mycoides* zählen auch *B. weihenstephanensis* und *B. wiedmannii* zu den psychrotoleranten Vertretern des Spezies Komplexes [166, 169]. Kältetolerante *B. cereus* s.l. Sporenbildner stellen als Verderbskeime ein zunehmendes Problem für die lebensmittelverarbeitende Industrie dar [202-205]. Zudem konnte die Existenz emetischer *B. weihenstephanensis* nachgewiesen werden [206, 207]. Im Gegensatz dazu handelt es sich bei *B. cytotoxicus* um eine thermotolerante Art mit einer eigenen Variante des Cytotoxins K (nämlich CytK-1) [167]. Der Typstamm der Spezies war verantwortlich für einen schwerwiegenden Enteritis Ausbruch mit drei Todesfällen in Frankreich [208]. Außerdem konnte CytK-1 im Kulturüberstand eines Endophtalmitis

verursachenden *B. cereus s.l.* Isolates ausfindig gemacht werden, wobei das Cytotoxin in der untersuchenden Studie als Hämolyisin IV (HlyIV) bezeichnet wurde [209]. Manche *B. cereus s.l.* Stämme sind aktive Bestandteile kommerziell vertriebener probiotischer Produkte [173, 210]. So wird der *B. toyonensis* Typstamm schon seit über 30 Jahren als Futtermittelzusatz (Toyocerin®) in der Nutztierwirtschaft (z.B. bei Schweinen, Geflügel oder Rindern) verwendet [168].

Mit Hilfe früher molekularbiologischer Techniken wie DDH, 16S rRNA Analyse oder Multilocus-Enzym-Elektrophorese (MEE) [211] war es nicht möglich die drei Arten *B. cereus*, *B. anthracis* und *B. thuringiensis* in ausreichendem Maße voneinander zu differenzieren [212-218]. Untersuchungen der *B. cereus s.l.* Populationsstruktur via Multilocus-Sequenz-Typisierung (MLST) [219] und amplifiziertem Fragmentlängen-Polymorphismus (AFLP) [220] ergaben, dass sich der Spezies Komplex in drei Kladen aufteilt. Klade I beinhaltet Isolate der Spezies *B. anthracis*, *B. cereus s.s.*, *B. thuringiensis* sowie emetische *B. cereus s.s.*, wohingegen Klade II nicht-emetische *B. cereus s.s.* wie auch *B. thuringiensis* Stämme umfasst. Klade III besteht aus den psychrotoleranten Arten *B. mycoides* und *B. weihenstephanensis* [184, 218, 221-226]. Einerseits bestätigten die MLST/AFLP Analysen, dass es sich bei *B. anthracis* um eine hoch klonale Abstammungslinie handelt [227-229]. Andererseits erbrachten sie den Beweis, dass *B. cereus s.s.* und *B. thuringiensis* genotypisch nicht voneinander zu unterscheiden sind, da die Stämme beider Arten in Klade I und Klade II vermischt vorliegen [184, 218, 221-226]. Dies ist nicht weiter verwunderlich, denn meistens liegen die  $\delta$ -Endotoxin Gene in *B. thuringiensis* Isolaten plasmidkodiert vor [230, 231]. Neben dem Nachweis der Produktion von Enterotoxinen [232-234] sind zudem Fälle von *B. thuringiensis* induzierten Lebensmittelerkrankungen (Diarrhoe Syndrom) und anderweitigen Infektionen des Menschen bekannt, die denen von *B. cereus s.s.* ähneln [183, 235-238]. Wie im Falle von *B. thuringiensis* liegen die Hauptvirulenzfaktoren des klonalen *B. anthracis* Komplexes auf extrachromosomalen Elementen. Das pXO1 Plasmid enthält die genetischen Determinanten für das Anthrax-Toxin, wohingegen das pXO2 Plasmid die Synthesegene für die antiphagozytäre Poly- $\gamma$ -D-Glutaminsäure-Kapsel beherbergt [239, 240]. Bekannt sind jedoch auch *B. cereus s.l.* Stämme, die Anthrax-ähnliche Erkrankungen hervorrufen, ohne zum klonalen Komplex zu gehören [241-244]. Charakteristisch für den klonalen Komplex ist eine Nonsense-Mutation im *plcR* Gen [240, 245, 246]. Der Phospholipase C Regulator (PlcR) ist der bedeutendste transkriptionelle Aktivator bekannter *B. cereus s.l.* Virulenzfaktoren [245, 247]. Ähnlich dem *B. anthracis*

Komplex sind emetische *B. cereus s.l.* auf klonale Abstammungslinien beschränkt [225, 226, 248-250], wobei das Cereulid-Synthetase-Gencluster in emetischen *B. weihenstephanensis* nicht nur plasmidischen sondern auch chromosomalen Ursprungs sein kann [251].

Nach dem Klassifikationsschema von Guinebrière *et al.* verteilen sich die neun valide beschriebenen *B. cereus s.l.* Typstämme auf sieben große phylogenetische Cluster: (I) *B. pseudomycooides*, (II) *B. wiedmannii*, (III) klonaler *B. anthracis* Komplex, (IV) *B. cereus s.s.* sowie *B. thuringiensis*, (V) *B. toyonensis*, (VI) *B. mycooides* sowie *B. weihenstephanensis* und (VII) *B. cytotoxicus* [252, 253]. Ursprünglich war das Klassifikationsschema auf den genetischen (fluorescent AFLP, 16S rRNA sowie *panC* Gensequenzanalyse) und phänotypischen Eigenschaften (Psychrotoleranz sowie Wachstum in unterschiedlichen Temperaturbereichen) von 425 *B. cereus s.l.* Isolaten begründet [252]. Später konnten Unterschiede im Ausmaß der Enterotoxizität/Zytotoxizität zwischen den phylogenetischen Clustern ausgemacht werden. So gelten die phylogenetischen Gruppen I, II und VII als hoch enterotoxisch, während Gruppe V moderat und Gruppe VI nur gering zytotoxisch ist. In Gruppen III und IV sind je nach Abstammungslinie alle drei Kategorien vertreten [160, 254, 255]. Derartige Zytotoxizitäts-Profile sowie die weitläufige Verbreitung von Enterotoxingenen innerhalb von *B. cereus s.l.* [256] liefern eine Erklärung dafür, dass die Fähigkeit das Diarrhoe Syndrom auszulösen nicht allein auf die pathogenen Vertreter der Spezies *B. cereus s.s.* beschränkt ist [237].

#### 4.2. Virulenzfaktoren, ihre genetischen Determinanten und Funktionen

Das Protektive Antigen (PA), der Letalfaktor (LF) und der Ödemfaktor (EF) repräsentieren das Anthrax-Toxin [257] und werden durch die drei Gene *pagA*, *lef* und *cya* kodiert [258-260]. Jedes der drei Exoproteine ist für sich genommen nicht toxisch. In Kombination bilden sie jedoch zwei binäre Toxine: das Letaltoxin (LT = PA + LF) und das Ödemtoxin (ET = PA + EF) [261]. Als Bindeeinheit ermöglicht PA nach erfolgreicher rezeptorvermittelter Endozytose die Translokation der beiden anderen Faktoren in das Zytoplasma der Wirtszelle [262]. Dort fungiert LF als Zn<sup>2+</sup>-abhängige Metalloprotease, die bestimmte MAP-Kinase Kinasen spaltet und somit inaktiviert [263-265], wohingegen EF als hocheffiziente Ca<sup>2+</sup>/Calmodulin-abhängige Adenylatzyklase die Umwandlung von ATP zu cAMP katalysiert [266-268]. In der frühen Phase einer *B. anthracis* Infektion stellt die toxische Wirkung der beiden Komponenten EF und LF gegenüber Zellen des angeborenen Immunsystems (wie

z.B. Makrophagen und neutrophilen Granulozyten) eine Ausweitung der Besiedelung des Wirtes hin zur systemischen Infektion sicher. Im Endstadium der Erkrankung führt ihre Aktivität dahingegen zum Tod des Wirtes aufgrund der Schädigung des Herz-Kreislauf-Systems sowie der Leber [269, 270].

Zu den Bt-Toxinen gehören die Cry (Crystal), Cyt (Cytolytic), Vip (Vegetative Insecticidal Protein) und Sip (Secreted Insecticidal Protein) Proteine. Nach ihrer proteolytischen Aktivierung im Insekten Darm binden die Proteine an spezifische Zelloberflächenrezeptoren und entfalten ihre toxische Wirkung, was zur Zerstörung des Mitteldarmepithels, Darmlähmung und dem Tod des Insekts führt [178, 271]. Die Benennung der Bt-Toxine erfolgt anhand der Sequenzähnlichkeit zu den bereits bekannten Cry, Cyt, Vip und Sip Proteinen nach einem Klassifikationsschema bestehend aus vier Rängen. Sequenzen mit einer Ähnlichkeit  $\geq 95\%$  unterscheiden sich lediglich im quartär-Rang. Sofern bei einem neuen Bt-Toxin dieser Ähnlichkeitsgrad (im Vergleich zu den bereits etablierten Proteinsequenzen) nicht erreicht werden kann, wird ihm ein neuer tertiär-Rang mit quartär-Rang '1' zugewiesen. Eine Sequenz mit quartär-Rang '1' wird als Holotyp bezeichnet [272]. Im Februar 2016 umfasste die Website des Komitees für die Nomenklatur von *B. thuringiensis*  $\delta$ -Endotoxinen (<http://www.btnomenclature.info/>) insgesamt 347 Toxin-Holotypen. Mit 304 Holotypen ist die Klasse der Cry Toxine am umfangreichsten und wird gefolgt von 32 Vip und elf Cyt Holotypen. Die bislang einzig bekannte Sip Holotyp-Sequenz Sip1Aa [273] wurde vom Komitee (noch) nicht offiziell aufgenommen. Das Wirtsspektrum der Bt-Toxine ist nicht allein auf Insekten (wie z.B. Coleoptera, Diptera, Lepidoptera, etc.) beschränkt. So gibt es auch Bt-Toxine, denen antibakterielle, nematizide oder akarizide Effekte nachgewiesen werden konnten. Darüber hinaus weist eine Untergruppe der Cry Toxine (die Parasporine) zellzerstörende Wirkung gegenüber bestimmten Tumorzelllinien des Menschen auf [178, 179].

Emetische *B. cereus* s.l. sind zur Biosynthese des Toxins Cereulid befähigt. Cereulid ist ein ringförmiges Dodekadepeptid mit der chemischen Struktur [D-O-Leu-D-Ala-L-O-Val-L-Val]<sub>3</sub> [182, 274]. Seine molekulare Zusammensetzung macht das Gift hitze-, säure- und alkalibeständig. Überdies kann es weder durch Trypsin oder Pepsin abgebaut werden [275-278]. Einmal im Lebensmittel gebildet und akkumuliert übersteht das Gift die Magen-Darmpassage [279]. Durch Fütterungsversuche mit Moschusspitzmäusen (*Suncus murinus*) konnte die Bindung des Toxins an die 5-

HT<sub>3</sub>-Rezeptoren des Duodenums nachgewiesen werden, was zur Vagusnerv-Stimulation mit anschließendem Brechreiz führte [280]. Die Strukturähnlichkeit von Cereulid zu Valinomycin lässt auf eine antibakterielle und antifungale Aktivität des Toxins schließen [281, 282], was emetischen *B. cereus s.l.* einen adaptiven Vorteil in ihrer ökologischen Nische (vermutlich Wurzeln und Wurzelknollen) verleihen könnte [250]. Außerdem liegt nahe, dass Cereulid als Ionophor [283] dazu dient dem Kalium-armem endophytischem und Boden-assoziiertem Milieu K<sup>+</sup>-Ionen zu entziehen [250, 284]. Die Bildung von Cereulid erfolgt durch eine nichtribosomale Peptidsynthetase (NRPS) [285, 286]. Das Cereulid-Synthetase-Gencluster *cesHPTABCD* umfasst sieben proteinkodierende Gene, wobei *cesA* und *cesB* die Strukturgene der NRPS repräsentieren. Die NRPS Module *CesA* und *CesB* sorgen für die Aktivierung und den Einbau von D-O-Leu sowie D-Ala (*CesA*) bzw. L-O-Val und L-Val (*CesB*). Bei *CesC* und *CesD* handelt es sich um einen putativen Transporter aus der ABC-Familie. Die genaue Funktion der vermeintlichen Hydrolase *CesH* ist noch nicht bekannt [287]. *CesP*, eine 4'-Phosphopantetheinyl Transferase, und *CesT*, eine Typ II Thioesterase, sind wahrscheinlich für die Aktivierung der NRPS (*CesP*) [288] respektive ihrer Regenerierung bei Einbau falscher Monomere (*CesT*) verantwortlich [289]. Die Regulation der Cereulid-Synthese ist abhängig von der Wachstumsphase des Produzenten, der Temperatur, der Verfügbarkeit von Sauerstoff, wie auch der Beschaffenheit des Lebensmittels/Wachstumsmediums [250, 256]. Der globale Regulator *CodY* [290] sowie der Zustandsübergangsregulator *AbrB* [291] stellen die zwei wichtigsten transkriptionellen Repressoren des Genclusters dar.

Die Diarrhoe Verlaufsform von *B. cereus s.l.* verursachten Lebensmittelerkrankungen wird durch Enterotoxin produzierende vegetative Zellen im Dünndarm hervorgerufen [160, 292]. Zu den Hauptvirulenzfaktoren in diesem Kontext zählen das nicht-hämolytische Enterotoxin (*Nhe*) [293], Hämolyysin BL (*HBL*) [294] und Cytotoxin K (*CytK*) [208]. *Nhe* und *HBL* sind zueinander homologe porenbildende Enterotoxin-Komplexe bestehend aus jeweils drei Proteinkomponenten [183, 256]. Die Gene beider Komplexe sind in den Operons *nheABC* und *hblCDAB* zusammengefasst [295-297]. Im Falle von *Nhe* entspricht *NheA* der lytischen Komponente, wohingegen *NheB* und *NheC* die Bindeproteine darstellen [297, 298]. Im Vero-Zell Modell zeigte der Komplex die maximale zytolytische Aktivität bei einem molaren Verhältnis von 10:10:1 (*NheA*:*NheB*:*NheC*) [299]. Ein inverted Repeat im *Nhe* Operon deutet auf die translationale Repression des *nheC* Genes hin, was dessen verhältnismäßig

geringeren Expressionsgrad im Vergleich zu den anderen beiden Genen erklärt [297, 299]. *NheABC* unterliegt einer strikt vertikalen Vererbung [26] und kommt in fast allen *B. cereus s.l.* vor [26, 254, 300, 301]. Außerdem existiert eine seltene homologe Variante *nheABC<sub>a</sub>* [26, 183, 302]. Die drei Untereinheiten L<sub>2</sub> (HbIC), L<sub>1</sub> (HbID) und B (HbIA) bilden den HBL Komplex, wobei die ersten beiden Proteine (L<sub>2</sub>, L<sub>1</sub>) eine lytische Funktion aufweisen und die B Untereinheit als Bindekomponente angesehen wird [294, 303]. Im Gegensatz zu *Nhe* zeigt HBL seine größte zytotoxische Wirkung bei einem äquimolaren Verhältnis seiner drei Proteinbestandteile [304]. Wahrscheinlich handelt es sich bei *hblB* um ein Pseudogen [183], das aufgrund einer *hblA* Duplikation hervorgegangen ist [305]. Wie schon im Falle von *Nhe<sub>a</sub>* kommt eine homologe Variante HBL<sub>a</sub> vor. Allerdings fehlt dem zugehörigen Operon *hblCDA<sub>a</sub>* das *hblB* Pseudogen [26, 183, 302, 306]. Die Transkription der *Nhe* und HBL Gene wird durch PlcR, dem Redox-Sensitive Two Component Signal Transduction System (ResDE) [307, 308] sowie dem Fumerate and Nitrate Reduction Regulator (Fnr) [309, 310] positiv beeinflusst, wohingegen das Catabolite Control Protein (CcpA) [311] als transkriptioneller Repressor fungiert [256, 292]. Das Einzelkomponenten Protein CytK ist ein Mitglied der Familie oligomerer  $\beta$ -Barrel porenbildender Toxine [208] von dem zwei Ausprägungen CytK-1 und CytK-2 mit einer Sequenzähnlichkeit von ~ 89 % in der *B. cereus s.l.* Gruppe existieren [312]. Überdies konnte CytK-1 gegenüber CytK-2 eine um den Faktor 5 erhöhte Zytotoxizität gegenüber Caco-2- und Vero-Zellen nachgewiesen werden [312]. Allerdings erscheint fraglich, ob die alleinige Existenz der CytK-1 Variante in einem Stamm eine hinreichende Bedingung für dessen gesteigertes pathogenes Potential darstellt. Während der *B. cytotoxicus* Typstamm NVH391/98<sup>T</sup> aufgrund seiner verstärkten *cytK-1* Expression hoch zytotoxisch gegenüber Vero-Zellen ist, handelt es sich beim Stamm NVH 883/00 um einen nicht toxischen Vertreter mit geringem *cytK-1* Expressionsgrad [313]. Beide CytK Ausprägungen stehen unter der transkriptionellen Kontrolle von PlcR [247, 256, 314]. Bisher konnte CytK-1 nur in Stämmen der Art *B. cytotoxicus* vorgefunden werden, womit CytK-2 das weiter verbreitete Toxin zu sein scheint [26, 254, 315]. Die Sekretion von *Nhe*, HBL wie auch CytK wird über das Sec-Translokationssystem gewährleistet [316].

Hämolysin II (HlyII) ist ein weiteres  $\beta$ -Barrel porenbildendes Toxin, das in der *B. cereus s.l.* Gruppe vorkommt und eine Sequenzähnlichkeit von ~ 37 % gegenüber CytK aufweist [208, 317, 318]. Die Expression des *hlyII* Genes ist im Gegensatz zum Großteil der *B. cereus s.l.* Virulenzfaktoren PlcR-unabhängig und wird durch den

Eisenaufnahme-Regulator (Fur) sowie den Hämolyisin II Repressor (HlyIIR) transkriptionell reguliert [319-322]. Eine Beteiligung von HlyII am Diarrhoe Syndrom erscheint aufgrund einer Trypsin Schnittstelle im Protein sowie der Akkumulation des Toxins unter Bedingungen mit hohen Oxidations-Reduktionspotentialen eher unwahrscheinlich [183, 208, 314, 323]. Das Toxin ist vielmehr Teil eines Mechanismus zur Abwehr der Immunantwort des Wirtes [314, 324]. Beispielsweise ruft die porenbildende Aktivität von HlyII eine Induktion der Apoptose bei Maus Makrophagen sowie menschlichen Monozyten hervor [325]. Neben Nhe, HBL, CytK und HlyII gehören auch das Cholesterol-abhängige Hämolyisin I (HlyI) alias Cereolysin O (CerO) [314, 326, 327] sowie das kaum untersuchte Hämolyisin III (HlyIII) [328] zum Repertoire der porenbildenden *B. cereus s.l.* Hämolyisine.

Des Weiteren verfügt *B. cereus s.l.* über drei Metalloproteasen, die als Immuninhibitoren A1, A2 und A3 (InhA1-3) bezeichnet werden [329-331]. InhA1 spielt eine zentrale Rolle bei der Makrophagen Evasion von *B. cereus s.l.* Sporen und vegetativen Zellen [332]. Zudem spaltet die Protease die beiden antimikrobiellen Peptide Attacin und Cecropin, die im Zuge der Pathogenabwehr in der Hämolymphe von Insekten vorhanden sind [329]. Überdies konnte gezeigt werden, dass InhA1 zur effizienten Hydrolyse von Gewebestandteilen wie Fibronectin und Collagen befähigt ist [333]. Zur genauen Rolle von InhA2 und InhA3 im Rahmen einer *B. cereus s.l.* Infektion ist nur wenig bekannt. Allerdings ist anzunehmen, dass die drei Metalloproteasen gleichzeitig beim Entgegenwirken der Immunabwehr sowie der Überwindung von Epithelbarrieren agieren [330, 331].

Mit den bisher aufgezählten Virulenzfaktoren ist das Arsenal an Enzymen, die zur Pathogenität von *B. cereus s.l.* beitragen, noch nicht erschöpft. So verfügen die Mitglieder der Gruppe mitunter über das Enterotoxin FM (EntFM) [334-336], Hämolyisin A (HlyA) sowie die drei Phospholipasen PI-PLC (Phosphatidylinositol-spezifische Phospholipase C), PC-PLC (Phosphatidylcholin-spezifische Phospholipase C) und SM-PLC (Sphingomyelin-spezifische Phospholipase C) [337-341].

## 5. Zielsetzung der Arbeit

Diese Arbeit verfolgt drei Zielsetzungen:

Zunächst soll das Illumina TruSeq® DNA PCR-free Sample Preparation Protokoll dahingehend modifiziert und optimiert werden, damit im Rahmen von Genomsequenzierungsprojekten paired-end Bibliotheken mit prolongierten Insert-Längen routinemäßig hergestellt werden können. Da allgemein bekannt ist, dass bei den NGS/TGS Technologien die Sequenzierung langer DNA-Fragmente mit einer Verschlechterung der Sequenzierqualität einhergeht, sollen derartige Abhängigkeiten auch bei den modifizierten Protokollen für bakterielle Genome mit unterschiedlichen GC-Gehalten genauer untersucht werden. Um bei prokaryotischen Sequenzierungsprojekten bestmögliche Genomrekonstruktionen garantieren zu können, sollte außerdem der Einfluss der Insert-Länge, der Read-Länge sowie der Sequenziertiefe auf die Leistungsfähigkeit von vier häufig genutzten Assemblern abgeschätzt werden. Auf dieser Grundlage sollen zahlreiche *Pseudomonas* und *B. cereus s.l.* Genome sequenziert werden.

Die genomischen Ähnlichkeitsmaße ANI<sub>b</sub> und ANI<sub>m</sub> gewinnen als neuer *in silico* Goldstandard zur Abgrenzung prokaryotischer Arten immer mehr an Bedeutung. In der Literatur werden bisher allerdings nur unscharfe Speziesgrenzen festgesetzt, weshalb in dieser Arbeit klar definierte Spezies-Grenzwerte für beide Ähnlichkeitsmaße abgeleitet werden sollen.

Schließlich sollen mit Hilfe geeigneter MLSA-Schemata und genomischen Speziesgrenzen die Arten der Gattung *Pseudomonas* und der *B. cereus s.l.* Gruppe analysiert werden. Neben den selbst sequenzierten Genomen sollten Genomsequenzen vom NCBI einbezogen werden. Genomospezies-Cluster, die auf Grundlage von *in silico* DDH Prozeduren und unter Anwendung strikter Spezies-Grenzwerte resultieren, sollen auf beide Phylogenien abgebildet und es soll überprüft werden, inwieweit sich die *Pseudomonas* bzw. *B. cereus s.l.* (Nomenspezies) Taxonomie mit den vorhergesagten Genomospezies-Clustern deckt. Darüber hinaus soll die Verteilung ausgewählter Virulenzfaktoren in der *B. cereus s.l.* Gruppe untersucht und geklärt werden, inwieweit sich Virulenzfaktoren als Spezies Marker eignen.

## IV. MATERIAL UND METHODEN

### 1. Chemikalien, Lösungen und Puffer

#### Chemikalie

L(+)-Ascorbinsäure  
 Hexadecyltrimethylammoniumbromid (CTAB)  
 Ethylendiamin-tetraessigsäure (EDTA) Dinatriumsalz Dihydrat  
 Glycerin  
 Lactose Monohydrat  
 Natriumacetat  
 Natriumchlorid (NaCl)  
 Natriumlaurylsulfat (SDS)  
 Natrium-L-Glutamat Monohydrat  
 Tris(hydroxymethyl)-aminomethan (TRIS)

#### Anbieter

Merck Millipore  
 Sigma-Aldrich  
 Carl Roth  
 Carl Roth  
 Merck Millipore  
 Carl Roth  
 Carl Roth  
 Carl Roth  
 PanReac AppliChem  
 Carl Roth

#### Lösung

2-Propanol  
 Ethanol  
 Natronlauge (NaOH)  
 Roti®-Phenol  
 Roti®-Phenol/Chloroform/Isoamylalkohol  
 Salzsäure (HCl)

Carl Roth  
 J.T.Baker  
 Carl Roth  
 Carl Roth  
 Carl Roth  
 Merck Millipore

#### Puffer

TE-Puffer

10 mM TRIS  
 1 mM EDTA Dinatriumsalz Dihydrat  
 pH: 8,0

Bis auf Hexadecyltrimethylammoniumbromid (CTAB) wurden sämtliche Chemikalien und Puffer in  $\text{H}_2\text{O}_{\text{dest}}$  gelöst. CTAB wurde in 0,7 M NaCl gelöst. Falls notwendig fand eine Einstellung des pH Wertes mit 1 M HCl bzw. 1 M NaOH statt. Die Sterilisation der Puffer und der in Lösung gegangenen Chemikalien fand bei 121 °C für 15 min im Autoklaven oder per Filtration (MILLEX® GP Filter; Ø 0,22 µm; Merck Millipore) statt. Phenolhaltige Lösungen wurden bei 4 °C gelagert. Ansonsten fand die Lagerung bei RT statt.

## 2. Medien

### Caseinpepton-Sojamehlpepton-Agar (Carl Roth)

15,0 g/l Pepton aus Casein  
5,0 g/l Pepton aus Sojamehl  
5,0 g/l Natriumchlorid  
15,0 g/l Agar-Agar

### Caseinpepton-Sojamehlpepton- Bouillon (Merck Millipore)

17,0 g/l Pepton aus Casein  
3,0 g/l Pepton aus Sojamehl  
2,5 g/l D(+)-Glucose Monohydrat  
5,0 g/l Natriumchlorid  
2,5 g/l di-Kaliumhydrogenphosphat

### Glycerin-Suspensionsmedium

10,0 g/l Natrium-L-Glutamat Monohydrat  
16,0 g/l Lactose Monohydrat  
1,0 g/l Agar-Agar (Oxoid)  
0,1 g/l L(+)-Ascorbinsäure  
120,0 g/l Glycerin

Caseinpepton-Sojamehlpepton-Agar (CASO-Agar) sowie Caseinpepton-Sojamehlpepton-Bouillon (CASO-Bouillon) wurden als Fertiggemische von der Carl Roth GmbH (Agar) und der Merck Millipore KGaA (Bouillon) kommerziell erworben. Die Komponenten/Fertiggemische sämtlicher Medien wurden in H<sub>2</sub>O<sub>dest.</sub> gelöst. Falls notwendig wurde der pH Wert mit 1 M HCl bzw. 1 M NaOH eingestellt. Die Sterilisation fand im Autoklaven bei 121 °C für 15 min statt. Gelagert wurde im Allgemeinen bei RT mit Ausnahme des Glycerin-Suspensionsmediums, welches bei 4 °C gelagert wurde.

## 3. Verwendete bakterielle Stämme und NCBI Daten

### 3.1. Verwendete bakterielle Stämme

#### 3.1.1. Bakterielle Stämme für die Optimierung der PCR-freien Probenvorbereitung

Die im Rahmen der Optimierung des Illumina TruSeq® DNA PCR-free Library Preparation Protokolls (Revision A, Januar 2013, Probenvorbereitung mit einer Insert-Länge von 550 bp) verwendeten bakteriellen Stämme sind in Tabelle 1 mit ihren Abkürzungen aufgelistet. Grundsätzlich wurden für jeden Stamm mehrere Sequenzierbibliotheken, die sich in ihrer Insert-Länge (Kategorie) unterschieden,

hergestellt. Welche Sequenzierbibliotheken (Kategorien) für welchen Stamm im Detail hergestellt und sequenziert wurden ist in Tabelle 3 (Seite 38) dargestellt.

**Tabelle 1** Verwendete bakterielle Stämme (Optimierung Probenvorbereitung)

Stammbezeichner	Abkürzung	RefSeq Sequenzen	% GC
<i>Bacillus cereus</i> F837/76	Bce	NC_016779.1, NC_016794.1, NC_016780.1	35,4
<i>Enterococcus faecalis</i> OG1RF	Efa	NC_017316.1	37,8
<i>Salmonella</i> Typhimurium 14028S	Sen	NC_016856.1, NC_016855.1	52,2
<i>Pseudomonas stutzeri</i> ATCC 17588 (T)	Pst	NC_015740.1	63,9
<i>Micrococcus luteus</i> NCTC 2665 (T)	Mlu	NC_012803.1	73,0

(T), anerkannter Typstamm nach LPSN [67]

Diese Tabelle wurde ins Deutsche übersetzt und entspricht Tabelle 1 aus der Studie von Huptas *et al.* [342].

Die Tabelle unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

### 3.1.2. Bakterielle Stämme für die Speziesabgrenzung

Im Zuge der Speziesabgrenzung mit Hilfe paarweiser Genomvergleiche (gANI, ANIm, ANIb) wurden acht Stämme der *Bacillus cereus* s.l. Gruppe, sowie 100 *Pseudomonas* Stämme sequenziert. Alle 108 selbst-sequenzierten Stämme sind in Tabelle 2 aufgelistet.

Im Falle der Gattung *Pseudomonas* wurden 30 Stämme (davon 29 anerkannte Typstämme) von der Deutschen Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ) bezogen. Die restlichen 70 Stämme stammten aus der Weihenstephaner Mikroorganismen-Sammlung (WS). Der offizielle Stammbezeichner (Gattungs- und Artnamen sowie Stammnummer) wurde bei den von der DSMZ bezogenen Stämmen übernommen. Im Hinblick auf die *Pseudomonas* Stämme aus der Weihenstephaner Mikroorganismen-Sammlung basiert die Zuordnung zu einer Art (siehe Stammbezeichner in Tabelle 2) auf der phylogenetischen Einzelgenanalyse des *rpoD*-Markergens zu den jeweiligen Typstammsequenzen. Für acht der 70 Stämme war die Zuordnung zu einer Art aufgrund fehlender Ähnlichkeit (< 97 % Sequenzähnlichkeit zur ähnlichsten Typstammsequenz) nicht eindeutig möglich. Sechs der acht Stämme waren Teil von zwei Neubeschreibungen mit persönlicher Beteiligung [343, 344]. Deren Vertreter sind entsprechend der publizierten Artnamen (*P. helleri*, *P. lactis*, *P. paralactis* und *P. weihenstephanensis*) benannt.

Sechs der acht Stämme aus der *B. cereus* s.l. Gruppe wurden vom Chemischen und Veterinäruntersuchungsamt Stuttgart (CVUAS) bezogen, die verbleibenden zwei

Stämme von der DSMZ. Für jeden Stamm wurde der offizielle Stammbezeichner (Gattungs- und Arname sowie Stammnummer) der entsprechenden Sammlung übernommen.

**Tabelle 2 Stämme, die in dieser Arbeit sequenziert wurden**

ID	Stammbezeichner	WS-Nr.	G-Nr.	Bezug
B001	<i>B. cytotoxicus</i> CVUAS 10162	5143	9642	CVUAS
B002	<i>B. cytotoxicus</i> CVUAS 3828	5137	8698	CVUAS
B003	<i>B. cytotoxicus</i> CVUAS 6082	5138	8700	CVUAS
B004	<i>B. cytotoxicus</i> CVUAS 644	5134	8694	CVUAS
B005	<i>B. cytotoxicus</i> CVUAS 6661	5141	9640	CVUAS
B006	<i>B. cytotoxicus</i> CVUAS 6688	5140	9641	CVUAS
B007	<i>B. mycoides</i> DSM 307	5133	9623	DSMZ
B008	<i>B. pseudomycoides</i> CVUAS 7789	5132	9691	CVUAS
P001	<i>P. aeruginosa</i> DSM 50071 (T)	2062	1273	DSMZ
P002	<i>P. aeruginosa</i> WS 5022	5022	9034	WS
P003	<i>P. alcaligenes</i> WS 5013	5013	7996	WS
P004	<i>P. brenneri</i> DSM 15294 (T)	4666	8128	DSMZ
P005	<i>P. brenneri</i> WS 5091	5091	9111	WS
P006	<i>P. chengduensis</i> DSM 26382 (T)	5003	8999	DSMZ
P007	<i>P. chengduensis</i> WS 5019	5019	7597	WS
P008	<i>P. chlororaphis</i> (ssp. <i>chlororaphis</i> ) DSM 50083 (T)	1625	1276	DSMZ
P009	<i>P. chlororaphis</i> WS 5014	5014	2638	WS
P010	<i>P. deceptionensis</i> DSM 26521 (T)	4853	8684	DSMZ
P011	<i>P. extremaustralis</i> DSM 17835 (T)	4974	9152	DSMZ
P012	<i>P. fluorescens</i> DSM 50090 (T)	2330	1288	DSMZ
P013	<i>P. fluorescens</i> WS 5037	5037	9060	WS
P014	<i>P. fragi</i> WS 5025	5025	9037	WS
P015	<i>P. fragi</i> WS 5032	5032	9048	WS
P016	<i>P. fragi</i> WS 5045	5045	9068	WS
P017	<i>P. fragi</i> WS 5065	5065	9087	WS
P018	<i>P. fragi</i> WS 5087	5087	9107	WS
P019	<i>P. fragi</i> WS 5094	5094	9114	WS
P020	<i>P. fragi</i> WS 5102	5102	9120	WS
P021	<i>P. fragi</i> WS 5112	5112	9129	WS
P022	<i>P. fragi</i> WS 5120	5120	9368	WS
P023	<i>P. fragi</i> WS 5123	5123	6634	WS
P024	<i>P. fragi</i> WS 5124	5124	7463	WS
P025	<i>P. fragi</i> WS 5125	5125	7835	WS
P026	<i>P. gessardii</i> DSM 17152 (T)	4665	8127	DSMZ
P027	<i>P. gessardii</i> WS 5049	5049	9072	WS
P028	<i>P. grimontii</i> DSM 17515 (T)	4659	8122	DSMZ
P029	<i>P. helleri</i> WS 4917 (T)	4917	8926	WS
P030	<i>P. helleri</i> WS 4995	4995	8966	WS
P031	<i>P. koreensis</i> DSM 16610 (T)	5007	9001	DSMZ
P032	<i>P. koreensis</i> WS 5015	5015	3883	WS

ID	Stammbezeichner	WS-Nr.	G-Nr.	Bezug
P033	<i>P. "lactis"</i> WS 4992 (iT)	4992	8927	WS
P034	<i>P. libanensis</i> DSM 17149 (T)	4854	8685	DSMZ
P035	<i>P. lini</i> DSM 16768 (T)	4860	8692	DSMZ
P036	<i>P. lundensis</i> DSM 6252 (T)	4855	8686	DSMZ
P037	<i>P. lundensis</i> WS 5024	5024	9036	WS
P038	<i>P. lundensis</i> WS 5026	5026	9042	WS
P039	<i>P. lundensis</i> WS 5071	5071	9056	WS
P040	<i>P. lundensis</i> WS 5078	5078	9099	WS
P041	<i>P. lundensis</i> WS 5081	5081	9057	WS
P042	<i>P. lundensis</i> WS 5088	5088	9108	WS
P043	<i>P. lundensis</i> WS 5095	5095	9058	WS
P044	<i>P. lundensis</i> WS 5104	5104	9059	WS
P045	<i>P. lundensis</i> WS 5110	5110	9127	WS
P046	<i>P. lundensis</i> WS 5117	5117	9365	WS
P047	<i>P. lundensis</i> WS 5118	5118	9366	WS
P048	<i>P. lundensis</i> WS 5119	5119	9367	WS
P049	<i>P. lundensis</i> WS 5121	5121	2949	WS
P050	<i>P. mandelii</i> DSM 17967 (T)	5009	8996	DSMZ
P051	<i>P. mandelii</i> WS 5114	5114	9131	WS
P052	<i>P. marginalis</i> DSM 13124 (T)	4861	8693	DSMZ
P053	<i>P. nitroreducens</i> DSM 14399 (T)	5005	8997	DSMZ
P054	<i>P. nitroreducens</i> WS 5012	5012	3904	WS
P055	<i>P. orientalis</i> DSM 17489 (T)	4927	8783	DSMZ
P056	<i>P. oryzihabitans</i> DSM 6835 (T)	5004	9000	DSMZ
P057	<i>P. oryzihabitans</i> WS 5017	5017	4781	WS
P058	<i>P. panacis</i> DSM 18529 (T)	4662	8124	DSMZ
P059	<i>P. panacis</i> WS 4668	4668	8218	WS
P060	<i>P. "paralactis"</i> WS 4672 (iT)	4672	8217	WS
P061	<i>P. peli</i> DSM 17833 (T)	5008	8995	DSMZ
P062	<i>P. peli</i> WS 5011	5011	9012	WS
P063	<i>P. poae</i> DSM 14936 (T)	4856	8687	DSMZ
P064	<i>P. poae</i> WS 5103	5103	9121	WS
P065	<i>P. protegens</i> WS 5082	5082	9102	WS
P066	<i>P. proteolytica</i> DSM 15321 (T)	4664	8126	DSMZ
P067	<i>P. proteolytica</i> WS 5010	5010	9011	WS
P068	<i>P. proteolytica</i> WS 5021	5021	9033	WS
P069	<i>P. proteolytica</i> WS 5027	5027	9043	WS
P070	<i>P. proteolytica</i> WS 5031	5031	9047	WS
P071	<i>P. proteolytica</i> WS 5059	5059	9081	WS
P072	<i>P. proteolytica</i> WS 5060	5060	9082	WS
P073	<i>P. proteolytica</i> WS 5079	5079	9100	WS
P074	<i>P. proteolytica</i> WS 5086	5086	9106	WS
P075	<i>P. proteolytica</i> WS 5111	5111	9128	WS
P076	<i>P. proteolytica</i> WS 5126	5126	9369	WS
P077	<i>P. proteolytica</i> WS 5127	5127	9370	WS
P078	<i>P. proteolytica</i> WS 5128	5128	9371	WS

ID	Stammbezeichner	WS-Nr.	G-Nr.	Bezug
P079	<i>P. proteolytica</i> WS 5145	5145	3687	WS
P080	<i>P. proteolytica</i> WS 5146	5146	4092	WS
P081	<i>P. proteolytica</i> WS 5147	5147	4642	WS
P082	<i>P. proteolytica</i> WS 5148	5148	7498	WS
P083	<i>P. psychrophila</i> DSM 17535 (T)	4857	8688	DSMZ
P084	<i>P. psychrotolerans</i> DSM 15758 (T)	5006	8998	DSMZ
P085	<i>P. rhodesiae</i> DSM 14020 (T)	4658	8121	DSMZ
P086	<i>P. rhodesiae</i> WS 4669	4669	3159	WS
P087	<i>P. rhodesiae</i> WS 5107	5107	9124	WS
P088	<i>P. sp.</i> WS 4996	4996	8963	WS
P089	<i>P. sp.</i> WS 5051	5051	9074	WS
P090	<i>P. stutzeri</i> WS 5018	5018	6238	WS
P091	<i>P. synxantha</i> DSM 18928 (T)	4858	8689	DSMZ
P092	<i>P. taetrolens</i> DSM 21104 (T)	4859	8690	DSMZ
P093	<i>P. trivialis</i> DSM 14937 (T)	4852	8691	DSMZ
P094	<i>P. veronii</i> DSM 11331 (T)	4661	8123	DSMZ
P095	<i>P. veronii</i> DSM 16272	4660	8109	DSMZ
P096	<i>P. veronii</i> WS 4670	4670	3537	WS
P097	<i>P. veronii</i> WS 4671	4671	3734	WS
P098	<i>P. veronii</i> WS 5113	5113	9130	WS
P099	<i>P. weihenstephanensis</i> WS 4993 (T)	4993	8928	WS
P100	<i>P. weihenstephanensis</i> WS 4994	4994	8959	WS

**ID**, Kennung zur eindeutigen Identifizierung eines Stammes

**Stammbezeichner**, Offizielle Bezeichnung eines Stammes in der jeweiligen Sammlung

**WS-Nr.**, Offizielle WS Stammnummer

**G-Nr.**, Nummer der Glycerinkryokultur eines Stammes am Lehrstuhl für Mikrobielle Ökologie, TUM

**Bezug**, Bezug des Stammes aus der jeweiligen Sammlung

(T), anerkannter Typstamm nach LPSN [67] (Stand Apr. 2017)

(iT), nicht anerkannter Typstamm, da Artname (noch) nicht in LPSN [67] aufgenommen (Stand Apr. 2017)

## 3.2. Verwendete NCBI Genomdaten

### 3.2.1. Genomdaten für die Optimierung der PCR-freien Probenvorbereitung

Tabelle 1 enthält für jeden verwendeten bakteriellen Stamm die Kennungen der NCBI Referenzgenomsequenzen (RefSeq Sequenzen), die bei der Assembly-Validierung im Rahmen der Optimierung des Illumina TruSeq® DNA PCR-free Sample Preparation Protokolls verwendet wurden. Neben den Kennungen ist auch der prozentuale GC-Gehalt der NCBI Referenzgenome aufgelistet.

### 3.2.2. Genomdaten für die Spezies-Grenzwertableitung

Für die genomischen Ähnlichkeitsmaße ANIb (beide Implementierungen) und ANIm wurde die Wahrscheinlichkeit, dass zwei prokaryotische Stämme zur gleichen Art gehören, mit Hilfe logistischer Regression statistisch modelliert (Methoden-Abschnitt

12. ). Im Zuge dieser Modellierung wurden die Genomdaten von 97 prokaryotischen Stämmen (darunter 61 anerkannte Typstämme) aus 27 Gattungen über das National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/>) bezogen. Diese sind in Tabelle 27 (Seite 187) mit ihren dazugehörigen Whole Genome Shotgun (WGS, <https://www.ncbi.nlm.nih.gov/Traces/wgs/>) Projekt Kennungen bei nicht-geschlossenen Genomassemblierungen, oder deren NCBI Referenzgenomsequenzen (RefSeq Sequenzen) bei geschlossenen Genomassemblierungen aufgelistet. Sofern bei nicht-geschlossenen Genomassemblierungen WGS Scaffold Sequenzen bezogen wurden, ist dies mit dem Kürzel „WGS\_SCAFLD“ gekennzeichnet. Die Bezeichnung aller Stämme (Gattungs- und Artzugehörigkeit, sowie Stammkennung) wurde vom NCBI übernommen.

### 3.2.3. Genomdaten für die Speziesabgrenzung

Neben den 100 selbst-sequenzierten Stämmen (Tabelle 2) wurden im Februar 2015 noch 270 zusätzliche Genomassemblierungen der Gattung *Pseudomonas* vom NCBI bezogen. Dieser Datensatz wurde im April 2017, aufgrund der starken Zunahme sequenzierter prokaryotischer Typstammgenome durch das GEBA-Projekt (Genomic Encyclopedia of Bacteria and Archaea) [345-347], nochmals um 93 NCBI-Genomassemblierungen erweitert. Zur Sicherstellung eines gewissen Mindestmaßes an Sequenzqualität wurden beide Male ausschließlich Assemblierungen mit  $\leq 300$  Contigs berücksichtigt. Im späteren Verlauf wurden 22 der bezogenen *Pseudomonas* Genomassemblierungen wieder verworfen (Tabelle 29 auf Seite 199). Die verwendeten *Pseudomonas* Genomdaten sind in Tabelle 28 (Seite 190) mit ihren jeweiligen WGS Projekt Kennungen oder NCBI/GenBank Referenzgenomsequenzen aufgelistet. Vereinzelt wurden auch WGS Scaffold Sequenzen bezogen, dies ist mit dem Kürzel „WGS\_SCAFLD“ kenntlich gemacht. 115 der 341 Stämme sind anerkannte Typstämme. Im Falle der Typspezies *P. aeruginosa* [101, 111] ist bekannt, dass der größte Teil des Genomes einzelner Isolate einem konservierten Core-Genom entspricht [348-351]. Um den Gesamtaufwand bei der Berechnung der paarweisen Genomvergleiche zu reduzieren, wurden lediglich 32 geschlossene *P. aeruginosa* Genomassemblierungen über das NCBI bezogen. Grundsätzlich wurden die Stammbezeichnungen (Gattungs- und Arname, sowie die Stammkennung) vom NCBI übernommen. Ausnahmen bilden dabei die Stämme *P. "cerasi"* 58, *P. perfectomarina* ATCC 14405, *P. sp. (batumici)* UCM B-321, *P. sp. (capeferrum)*

WCS358, *P. sp. (defensor)* WCS374, *P. sp. (saudiphocaensis)* 20\_BN und *P. "versuta"* L10.10, die beim NCBI die Stammbezeichnungen *P. sp. 58*, *P. stutzeri* ATCC 14405, *P. batumici* UCM B-321, *P. capeferrum* WCS358, *P. sp.* WCS374, *P. sp.* 20\_BN und *P. sp.* L10.10 tragen.

Im Rahmen der Speziesabgrenzung innerhalb der *B. cereus s.l.* Gruppe wurden neben den acht selbst-sequenzierten Stämmen (Tabelle 2) zunächst noch 294 Genomassemblierungen über das NCBI bezogen (Oktober 2015). Aus Qualitätsgründen wurden generell nur Assemblierungen mit  $\leq 400$  Contigs/Scaffolds berücksichtigt. Genomdatensätze mit den Hinweisen „misassembled“, „partial“ oder „many frameshifted proteins“ fanden auch keine Berücksichtigung. Im Falle von *B. anthracis* wurden lediglich acht geschlossene Genomassemblierungen herangezogen, da es sich um eine monophyletische, hoch klonale Gruppe handelt [240, 352]. Von den 294 bezogenen Genomassemblierungen wurden 33 allerdings wieder verworfen. Nach einer ersten Berechnung aller paarweisen AF und gANI Werte war aufgefallen, dass einige Stämme zwar unterschiedliche Stammkennungen tragen ihre paarweisen AF und gANI Werte allerdings hoch ähnlich sind. In diesen Fällen wurde lediglich ein Referenzstamm ausgewählt und alle anderen verworfen. Zudem gibt es Stämme für die mehrere Genomdatensätze (Duplikate) beim NCBI vorhanden sind. Auch hier wurde stellvertretend nur ein Referenzdatensatz gewählt. Die zwei Stämme *B. thuringiensis* 1094BTHU und *B. mycoides* 11kri323 wurden verworfen, da ihre Genomassemblierungen einen zu niedrigen bzw. zu hohen prozentualen GC-Gehalt für die *B. cereus s.l.* Gruppe aufweisen. Neben den charakterisierten Stämmen des NCBI (Artnamen vorhanden) wurden auch Assemblierungen von 27 nicht-charakterisierten Stämmen (*B. sp.*), die einen GC-Gehalt von 34-37 % und eine Assembly Size  $\geq 4$  Mbp aufweisen, bezogen (Oktober 2015). Auf Grundlage eines Screenings nach *B. cereus s.l.* Mitgliedern unter den nicht-charakterisierten Stämmen wurden zwölf wieder verworfen, da ihre maximale AF zu einem beliebigen anderen Stamm der bisher 261 charakterisierten (nicht verworfenen) Stämme kleiner als 0,25 war. Nach dem MLSA Screening (Methoden-Abschnitt 10.2. ) wurden aufgrund unvollständiger oder fehlender Gene auch die Stämme *B. cereus* H3081.97, *B. cereus* K-5975c und *B. mycoides* Rock1-4 nicht weiter berücksichtigt. Somit wurden insgesamt 273 Genomassemblierungen über das NCBI bezogen und weiter verwendet. Diese sind in Tabelle 31 (Seite 202) mit ihren WGS Projekt Kennungen oder ihren RefSeq/GenBank Referenzgenomsequenzen zusammengefasst. Unter diesen 273 Stämmen befinden

sich sieben anerkannte Typstämme. Die drei Stämme *B. "bingmayongensis"* FJAT-13831 (*B. sp.* FJAT-13831), *B. "gaemokensis"* JCM 15801 und *B. "manliponensis"* JCM 15802 entsprechen den Typstämmen bereits beschriebener Arten [170-172], wobei die Artnamen noch nicht in die LPSN aufgenommen wurden. Deshalb werden die drei Stämme mit dem Kürzel „iT“ (invalid type strain) als nicht anerkannte Typstämme geführt. Alle verworfenen NCBI Genomassemblierungen sowie den Grund ihres Ausschlusses finden sich in Tabelle 32 (Seite 211). Grundsätzlich wurden sämtliche Stammbezeichnungen (Gattungs- und Arname, sowie die Stammkennung) vom NCBI übernommen.

### 3.3. Verwendete NCBI Sequenzdaten

#### 3.3.1. Sequenzdaten für die Multilocus-Sequenz-Analyse (Speziesabgrenzung)

Für die Gattung *Pseudomonas* wurde das weitläufig verwendete MLSA-Schema von Mulet *et al.* [25] herangezogen (Methoden-Abschnitt 10.1. ). Dabei wurden die Nukleotidsequenzen der Gene 16S rDNA, *gyrB*, *rpoB* und *rpoD* von 19 Typstammgenomen, für die keine Genomassemblierungen verfügbar waren, vom NCBI bezogen. Zudem kam es vor, dass für bestimmte *Pseudomonas* Stämme die entsprechenden Einzelgensequenzen unzureichend oder gar nicht in den zur Verfügung stehenden Genomassemblierungen gefunden werden konnten (Methoden-Abschnitt 10.2. ). Auch in diesen Fällen wurde auf NCBI-Einzelgensequenzen zurückgegriffen. Tabelle 30 (Seite 200) gibt Aufschluss über die betroffenen Stämme und verwendeten NCBI-Sequenzdaten.

## 4. Kultivierung und Lagerung bakterieller Stämme

Grundsätzlich wurden die verwendeten bakteriellen Stämme mit Hilfe eines Verdünnungsausstriches auf CASO-Agar bei 30 °C unter aeroben Bedingungen für 1-2 Tage (je nach bakterieller Gattung) inkubiert. Die kurzzeitige Lagerung (1-2 Wochen) der Kulturen fand bei 4 °C statt.

Für die langfristige Lagerung wurden Glycerinkryokulturen angelegt. Hierzu wurde ausgehend von einer koloniebildenden Einheit (KbE) ein Massenausstrich auf CASO-Agar angefertigt. Nach Bebrütung (30 °C, aerob, 1-2 Tage) wurde der Massenausstrich mit einer Impföse in ein Glycerin-Suspensionsmedium gemischt. Die Kryokulturen wurden zunächst bei 4 °C über Nacht inkubiert, bevor sie bei -80 °C gelagert wurden.

Übernachtskulturen (ÜNK) wurden ausgehend von jeweils einer KbE in 4 ml CASO-Bouillon bei 30 °C unter aeroben Bedingungen und kontinuierlichem Schütteln (150 rpm) angezogen.

## **5. DNA Isolierung und Konzentrationsbestimmung**

### **5.1. DNA Isolierung aus gram-positiven Bakterien**

Genomische DNA gram-positiver Bakterien wurde unter Verwendung von Hexadecyltrimethylammoniumbromid (CTAB) isoliert [353].

#### **5.1.1. Mechanischer Aufschluss**

Um die DNA Ausbeute bei gram-positiven Bakterien zu erhöhen wurden deren Zellen zunächst mechanisch aufgeschlossen. Hierzu wurden 3 ml einer bakteriellen ÜNK bei 15.700 g und RT für 2 min pelletiert. Das Zellpellet wurde anschließend in 567 µl TE-Puffer resuspendiert. Nach Zugabe von Zirconiumsilicat-Kügelchen (Ø 0,1 mm, BioSpec Products) und unter Verwendung eines MP FastPrep<sup>®</sup>-24 Instrumentes (MP Biomedicals) wurden die bakteriellen Zellen insgesamt zweimal bei 6,5 m/s für jeweils 45 s aufgeschlossen.

#### **5.1.2. DNA/RNA Isolierung**

Nach Abschluss des Zellaufschlusses wurden der Probe 30 µl einer 10 % (w/v) SDS-Lösung sowie 3 µl einer 20 mg/ml Proteinase K-Lösung (Macherey-Nagel) zugegeben. Nach kurzem Mischen erfolgte die Inkubation bei 37 °C für 3 h, um die durch den Zellaufschluss freiliegenden Proteine zu inaktivieren. Im Anschluss wurden 100 µl einer 5 M NaCl-Lösung sowie 80 µl bereits vorgewärmter 10 % (w/v) CTAB-Lösung (60 °C) zugegeben und mit einem Reagenzglasschüttler gemischt. Nach erfolgter Inkubation bei 65 °C für 35 min wurden 780 µl Roti<sup>®</sup>-Phenol zugegeben und vermischt (Reagenzglasschüttler). Die erste DNA/RNA Extraktion erfolgte nach Zentrifugation des Gemisches bei 15.700 g und RT für 4 min, indem die wässrige Phase vorsichtig in ein neues Mikroreaktionsgefäß überführt wurde. Analog der ersten DNA/RNA Extraktion schlossen sich noch zwei weitere an, wobei zuerst 700 µl und anschließend 600 µl Roti<sup>®</sup>-Phenol/Chloroform/Isoamylalkohol verwendet wurden. Nach der dritten DNA/RNA Extraktion wurden der wässrigen Phase 0,6 Volumen (250-300 µl) 2-Propanol zugegeben. Die daraufhin eintretende Fällung der DNA/RNA wurde für 15 min bei 4 °C fortgesetzt. Nach Zentrifugation bei 15.700 g und RT für 4 min, wurde das DNA/RNA-Pellet mit 70 % (v/v) Ethanol gewaschen,

bevor es bei 37 °C für 5 min getrocknet wurde. Abschließend wurde das DNA/RNA-Pellet in 100 µl H<sub>2</sub>O<sub>dest.</sub> gelöst und über Nacht bei 4 °C gelagert.

### 5.1.3. RNA Verdau

1 µl einer 10 mg/ml RNase A-Lösung (Thermo Scientific) wurden mit der in 100 µl H<sub>2</sub>O<sub>dest.</sub> gelösten DNA/RNA vermischt. Der RNA Verdau fand bei 37 °C für 30 min statt. Nach Zugabe von 300 µl H<sub>2</sub>O<sub>dest.</sub> sowie 400 µl Roti®-Phenol/Chloroform/Isoamylalkohol wurde die Probe gemischt (Reagenzglasschüttler). Das Gemisch wurde bei 15.700 g und RT für 4 min zentrifugiert und die wässrige Phase vorsichtig in ein neues Mikroreaktionsgefäß überführt. Nach Zugabe von 0,1 Volumen (30-40 µl) einer 3 M Natriumacetat-Lösung (pH 5,2) und 2,5 Volumen (750-1000 µl) eiskaltem 100 % Ethanol fand die DNA Fällung bei -20 °C für 20 min statt. Anschließend wurde das DNA-Pellet mit 70 % (v/v) Ethanol gewaschen und für 5 min bei 37 °C getrocknet, bevor es in 100 µl H<sub>2</sub>O<sub>dest.</sub> gelöst und über Nacht bei 4 °C inkubiert wurde. Dauerhaft gelagert wurden die DNA Proben bei -20 °C.

## 5.2. DNA Isolierung aus gram-negativen Bakterien

Die Isolierung genomischer DNA aus gram-negativen Bakterien fand mit Hilfe des QIAamp® DNA Mini Kits (Qiagen) unter Anwendung des Protokolls zur Isolierung genomischer DNA ausgehend von einer bakteriellen Einzelkultur auf Kulturplatte statt (QIAamp® DNA Mini and Blood Mini Handbook, Third Edition, June 2012, p. 56 and pp. 33-35). Entgegen den Herstellerangaben betrug die Inkubationszeit der Proteinase K Behandlung insgesamt 4 Stunden bei 56 °C. Der RNA-Verdau fand nach Zugabe von 10 µl einer 10 mg/ml RNase A-Lösung (Thermo Scientific) bei 70 °C für 30 min statt. Die genomische DNA wurde zweimal mit jeweils 100 µl H<sub>2</sub>O<sub>dest.</sub> eluiert. Dauerhaft gelagert wurden die DNA Proben bei -20 °C.

## 5.3. Bestimmung der DNA Konzentration

Die DNA Konzentrationen der einzelnen Proben wurden unter Verwendung eines Qubit® 2.0 Fluorometers (Invitrogen) und des Qubit™ dsDNA HS Assay Kits (Invitrogen) nach Herstellerangaben (Qubit™ dsDNA HS Assay Kits, MAN0002326, October 2010) bei ~ 21 °C RT bestimmt. Für jede angesetzte Working Solution erfolgte die Kalibrierung des Fluorometers erneut. Das verwendete Volumen der DNA Proben lag bei jeweils 1 µl. Jede Probe wurde als unabhängiges Duplikat oder

Triplikat mehrmals gemessen. Der sich daraus ergebende Durchschnittswert wurde als endgültiger DNA Konzentrationswert angesehen.

## 6. *De novo* Genomsequenzierung und Assemblierung

### 6.1. Optimierung der PCR-freien Probenvorbereitung

Das Illumina TruSeq® DNA PCR-free Sample Preparation Protokoll (Revision A, Januar 2013, Probenvorbereitung mit einer Insert-Länge von 550 bp) für die *de novo* Sequenzierung ganzer Genome wurde dahingehend optimiert, dass Modifikationen bei der DNA-Fragmentierung, der Größenselektion der DNA-Fragmente und der Konzentrationsbestimmung des fertigen Bibliothekenpools eingeführt wurden. Diese Modifikationen erlauben die Herstellung von Sequenzierbibliotheken, welche sich in ihrer durchschnittlichen Insert-Länge unterscheiden. Mit den beschriebenen Modifikationen können vier unterschiedliche Kategorien (Kategorien IS1-IS4) von Sequenzierbibliotheken hergestellt werden. Dabei gilt, dass Bibliotheken der Kategorie IS1 eine kürzere durchschnittliche Insert-Länge als Bibliotheken der Kategorie IS2 aufweisen und diese wiederum eine kürzere durchschnittliche Insert-Länge als Bibliotheken der Kategorie IS3 besitzen ( $IS1 < IS2 < IS3 < IS4$ ). Insgesamt wurden im Rahmen der Optimierung der Probenvorbereitung 5 bakterielle Stämme verwendet (Tabelle 1), deren genomischer GC-Gehalt sich von 35,4 % bis 73,0 % erstreckt. Tabelle 3 zeigt die Kategorien der Sequenzierbibliotheken, die für jeden Stamm hergestellt und sequenziert wurden. Dabei steht Kategorie TS für Bibliotheken, die nach Vorgabe des unveränderten Illumina TruSeq® DNA PCR-free Sample Preparation Protokolls (Revision A, Januar 2013, Probenvorbereitung mit einer Insert-Länge von 550 bp) vorbereitet und sequenziert wurden. Unabhängig zur Kategorie einer Sequenzierbibliothek wurden für ihre Herstellung ausschließlich die Puffer und Reagenzien des kommerziell erworbenen TruSeq® DNA PCR-free Sample Prep LT Kits (Illumina) verwendet.

Tabelle 3 Kategorien der Sequenzierbibliotheken

Kategorie	Fragmentierungseinstellungen	RB:BR Verhältnisse	Ø Insert-Länge (Rohdaten)	Sequenzierte Genome
TS	DF 5 %, PIP 175 W, C/B 200, Du 25 sec	TruSeq® DNA PCR-free Protokoll (550 bp)	641 ± 28 bp	Bce, Efa, Pst, Mlu
IS1	DF 10 %, PIP 175 W, C/B 200, Du 25 sec	2,0:1 + 3,8:2	686 ± 33 bp	Bce, Efa, Sen, Pst, Mlu
IS2	DF 2 %, PIP 175 W, C/B 200, Du 30 sec	2,2:1 + 4,2:2	990 ± 79 bp	Efa, Sen, Mlu <sub>50</sub>
IS3	DF 2 %, PIP 175 W, C/B 200, Du 20 sec	2,3:1 + 4,4:2	1.211 ± 78 bp	Bce, Pst, Mlu, Mlu <sub>50</sub>
IS4	DF 2 %, PIP 175 W, C/B 200, Du 10 sec	2,4:1 + 4,6:2	1.297 bp	Mlu <sub>50</sub>

TS, Original Illumina TruSeq® DNA PCR-free Protokoll (keine Modifikationen).

IS1-IS4, Kategorien der Sequenzierbibliotheken, wobei für die Insert-Länge gilt: IS1 < IS2 < IS3 < IS4.

Bce, *B. cereus* F837/76; Efa, *E. faecalis* OG1RF; Pst, *P. stutzeri* ATCC 17588<sup>T</sup>; Mlu, *M. luteus* NCTC 2665<sup>T</sup>; Sen, *S. Typhimurium* 14028S.

Die Bezeichnung Mlu<sub>50</sub> steht für die Sequenzierung des *M. luteus* NCTC 2665<sup>T</sup> Genomes bei 2x25 bp. Genome, die ohne Index gekennzeichnet sind, wurden durchgehend mit 2x200 bp sequenziert.

RB, Resuspension Buffer; BR, Magnetic Bead Reagent; DF, Duty Factor; PIP, Peak Incident Power; C/B, Cycles per Burst; Du, Duration.

Diese Tabelle wurde ins Deutsche übersetzt und entspricht Tabelle 2 aus der Studie von Huptas *et al.* [342].

Die Tabelle unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

### 6.1.1. Fragmentierung genomischer DNA

Die unveränderte Probenvorbereitung (Kategorie TS) startet mit 52,5 µl genomischer DNA gelöst in Resuspension Buffer (RB, Illumina) in einer Konzentration von 40 ng/µl. Im Gegensatz dazu liegt bei der modifizierten Probenvorbereitung (Kategorien IS1-IS4) die Ausgangskonzentration der genomischen DNA bei 75 ng/µl in 52,5 µl RB. Die Fragmentierung der genomischen DNA fand immer in microTUBEs (Covaris) und unter Verwendung des S220 Focused-ultrasonicator Systems (Covaris) bei einer Wassertemperatur von 6 °C im frequency sweeping Modus statt. Alle weiteren Kategorie-abhängigen Fragmentierungseinstellungen können Tabelle 3 entnommen werden.

### 6.1.2. Erste Längenselektion der DNA-Fragmente

Die Anwendung unterschiedlicher Volumenverhältnisse von Resuspension Buffer (RB, Illumina) zu Agencourt AMPure® XP Reagent (BR, Beckman Coulter) ermöglicht die Selektion von DNA-Fragmenten unterschiedlicher Länge. Die RB:BR Volumenverhältnisse sowie die zugrundeliegenden individuellen Volumina, welche

zur Herstellung der Sequenzierbibliotheken unterschiedlicher Kategorie (IS1-IS4) notwendig sind, sind in Tabelle 3 und Tabelle 35 (Seite 217) aufgeführt. Im Folgenden wird ausschließlich die erste Längenselektion bei der Vorbereitung von IS1 Sequenzierbibliotheken im Detail beschrieben.

Zu 50  $\mu$ l fragmentierter genomischer DNA wurden 30  $\mu$ l RB und 40  $\mu$ l BR hinzugefügt. Nach vorsichtigem Mischen durch wiederholtes Auf- und Abpipettieren wurde die Lösung für 5 min bei RT inkubiert. Zu diesem Zeitpunkt betrug das RB:BR Volumenverhältnis 80:40 ( $\cong$  2:1). Bei diesem Volumenverhältnis binden DNA Fragmente, welche zu lang für Sequenzierbibliotheken der Kategorie IS1 sind, an die magnetischen Beads. Nach Überführung der Probe auf einen Magnetständer (MagJET™ Separation Rack 12 x 1,5 ml, Thermo Scientific) und Inkubation für 3 min konzentrierten sich die magnetischen Beads als Pellet in der Lösung. Nun wurden 117  $\mu$ l des klaren (Bead-freien) Überstandes in ein neues Mikroreaktionsgefäß überführt. Anschließend wurden 70,2  $\mu$ l RB und 39  $\mu$ l 2X BR hinzugefügt. Die Angabe 2X entspricht hier der doppelten Konzentration magnetischer Beads pro Volumen des reinen (Bead-freien) BR Puffers. Die Verwendung einer höheren Beadkonzentration bei diesem Schritt ermöglicht das Binden einer höheren Menge von DNA-Fragmenten gewünschter Länge im späteren Verlauf. Durch wiederholtes Auf- und Abpipettieren wurde die Lösung vorsichtig gemischt, wobei das RB:BR Volumenverhältnis nun 148:78 ( $\cong$  3,8:2,0) betrug. Bei diesem Volumenverhältnis verbleiben zu kurze DNA Fragmente für eine Sequenzierbibliothek der Kategorie IS1 im Überstand. DNA Fragmente der gewünschten Länge binden an die magnetischen Beads. Das Gemisch wurde für 5 min inkubiert. Anschließend wurde die Probe für 3 min auf einen Magnetständer überführt. Nach Bildung des Bead-Pellets wurde der klare Überstand verworfen. Mit Verbleib der Probe auf dem Magnetständer wurde das Bead-Pellet zweimal mit jeweils 200  $\mu$ l frisch vorbereitetem 80 % (v/v) Ethanol vorsichtig gewaschen ohne es zu resuspendieren. Abschließend wurde das Bead-Pellet für 5 min luftgetrocknet, bevor die DNA-Fragmente gewünschter Länge von den magnetischen Beads in 52  $\mu$ l RB eluiert wurden.

### **6.1.3. End Repair**

Die Enden der längenselektierten DNA-Fragmente (Methoden-Abschnitt 6.1.2. ) liegen aufgrund der Fragmentierung der genomischen DNA via Ultraschallbehandlung (Methoden-Abschnitt 6.1.1. ) in drei Formen vor: mit glattem Ende (blunt end), mit 5'-Überhang (sticky end) oder mit 3'-Überhang (sticky end).

Enzyme mit 5'-3'-Polymeraseaktivität füllen 5'-Überhänge auf, wohingegen Enzyme mit 3'-5'-Exonukleaseaktivität 3'-Überhänge auflösen. Unter End Repair versteht man die Nutzung dieser enzymatischen Aktivitäten, um ausgehend von DNA-Fragmenten mit Überhängen DNA-Fragmente mit glatten Enden zu erhalten.

50 µl gröÙenselektierter DNA wurden 5 µl End Repair Control (Illumina) und 20 µl End Repair Mix 2 (Illumina) hinzugefügt. Die End Repair Reaktion fand unter Verwendung eines vorgewärmten QBT 2 Heizblockes (Grant Instruments) für 35 min bei 30 °C statt.

#### **6.1.4. Zweite Längenselektion der DNA-Fragmente**

Prinzipiell wurde die zweite Längenselektion entsprechend der ersten Längenselektion (Methoden-Abschnitt 6.1.2. ) durchgeführt. Die zweite Längenselektion startet im Gegensatz zur Ersten allerdings mit einem Probenvolumen von 70 µl anstatt 50 µl. Dementsprechend war darauf zu achten, dass einer Probe 10 µl RB (IS1), 18 µl RB (IS2), 22 µl RB (IS3) oder 26 µl RB (IS4) hinzugefügt werden musste, bevor mit Schritt 3 aus Tabelle 35 fortgefahren werden konnte. Am Ende der zweiten Längenselektion wurden die DNA-Fragmente gewünschter Länge von den magnetischen Beads in 19 µl RB eluiert.

#### **6.1.5. 3'-Adenylierung und Ligation der Adapter**

Während der Adenylierungsreaktion wird jeweils ein Adenin-Nukleotid an beiden 3'-Enden der glatten DNA-Fragmente angehängt. Dadurch soll die Bildung von DNA-Fragment-Konkatemeren (Chimären) während der Ligationsreaktion erschwert werden. Die Illumina TruSeq® Adapter besitzen am 3'-Ende ein überhängendes terminales Thymin-Nukleotid. Aufgrund des komplementären Überhanges zwischen DNA-Fragment und Adapter wird die Effizienz der Adapterligation erhöht.

Ausgehend von 15 µl gröÙenselektierter DNA wurden 1,25 µl A-Tailing Control (Illumina) und 6,25 µl A-Tailing Mix (Illumina) hinzugefügt. Die Adenylierungsreaktion wurde auf einem vorgewärmten QBT 2 Heizblock (Grant Instruments) bei 37 °C für 35 min ausgeführt. Dem folgte die Enzyminaktivierung bei 70 °C für 5 min. Abschließend wurde die Probe für 5 min auf Eis gelegt.

22,5 µl der Adenylierungsreaktion wurden 1,25 µl Ligation Control (Illumina), 1,25 µl Ligation Mix 2 (Illumina) und 1,25 µl der gewünschten DNA Adapter Lösung (Illumina) zugegeben. Die Ligationsreaktion fand auf einem vorgewärmten QBT 2

Heizblock bei 30 °C für 12 min statt. Daraufhin wurde das Enzym durch Zugabe von 2,5 µl Stop Ligation Buffer (Illumina) inaktiviert.

#### **6.1.6. Aufreinigung der fertigen Sequenzierbibliothek**

Bei der Aufreinigung der fertigen Sequenzierbibliothek soll die Probe von Kontaminanten, wie den Enzymen der Adenylierungs- und Ligationsreaktion, Adapter-Monomeren, Adapter-Dimeren und DNA-Fragment-Konkatemeren, befreit werden.

Ausgehend von 25 µl der Ligationsreaktion wurden 25 µl BR hinzugefügt, was einem RB:BR Volumenverhältnis von 1:1 entspricht. Die Probe wurde durch wiederholtes und vorsichtiges Auf- und Abpipettieren vermischt. Nach Inkubation des Gemisches für 5 min wurde die Probe auf einen Magnetständer (MagJET™ Separation Rack 12 x 1.5 ml, Thermo Scientific) überführt und es wurde 3 min gewartet bis sich das Bead-Pellet gebildet hatte. Daraufhin wurde der klare Überstand verworfen. Während die Probe auf dem Magnetständer verblieb, wurde das Bead-Pellet zweimal mit jeweils 200 µl frisch vorbereitetem 80 % (v/v) Ethanol gewaschen ohne es aufzulösen. Schließlich wurde das Bead-Pellet für 5 min luftgetrocknet, bevor die fertige Sequenzierbibliothek von den magnetischen Beads in 25 µl RB eluiert wurde.

Insgesamt fanden zwei Aufreinigungen statt. Am Ende der zweiten Aufreinigung wurde die Bibliothek in 15 µl RB (anstatt 25 µl RB) eluiert.

#### **6.1.7. Validierung und Quantifizierung der Sequenzierbibliothek**

Die Bestimmung der durchschnittlichen Insert-Länge einer Sequenzierbibliothek mit Hilfe des 2100 Bioanalyzer Instrumentes (Agilent Technologies) ist direkt nach der zweiten Größenselektion (vor der Adapterligation) erforderlich. Die Illumina TruSeq® Adapter besitzen eine gabelartige (forked) Struktur, da die beiden DNA-Stränge eines Adapters nicht über ihre gesamte Länge komplementär zueinander sind. Diese Gabelstruktur würde zu einem anormalen Migrationsverhalten von DNA-Fragmenten mit bereits ligierten Adapters bei der elektrophoretischen Auftrennung führen und damit zu einer fehlerhaften Bestimmung der durchschnittlichen Insert-Länge einer Sequenzierbibliothek.

Direkt nach der zweiten Längenselektion (Methoden-Abschnitt 6.1.4. ) wurde 1 µl der Probe für die DNA Konzentrationsmessung (Methoden-Abschnitt 5.3. ) verwendet. Auf eine Mehrfachmessung (unabhängige Duplikate) der Probe wurde dabei

verzichtet. Anschließend wurden 2 µl der Probe auf eine Endkonzentration von ~ 0,3 ng/µl verdünnt. Die durchschnittliche Insert-Länge der Sequenzierbibliothek wurde per Dreifachmessung bestimmt. Dabei wurde jeweils 1 µl der verdünnten (~ 0,3 ng/µl) Probe unter Verwendung des 2100 Bioanalyzer Instrumentes auf einem High Sensitivity DNA Chip (Agilent Technologies) nach Herstellerangaben (Agilent High Sensitivity DNA Kit Quick Start Guide, Revision C) elektrophoretisch aufgetrennt. Aus allen drei Messungen wurde der Durchschnitt gebildet. Der sich daraus ergebende Wert wurde als die mit Hilfe des Bioanalyzers ermittelte durchschnittliche Insert-Länge der Sequenzierbibliothek ( $AIS_{Bio}$ ) definiert.

Nach Aufreinigung der fertigen Sequenzierbibliotheken (Methoden-Abschnitt 6.1.6. ) wurde die DNA Konzentration jeder Bibliothek in ng/µl ( $conc$ ), wie in Methoden-Abschnitt 5.3. beschrieben, bestimmt. Die Molarität der ersten fertigen Sequenzierbibliotheken wurde dadurch abgeschätzt, dass die durch den Bioanalyzer ermittelte durchschnittliche Insert-Länge in bp ( $AIS_{Bio}$ ) als die tatsächliche durchschnittliche Insert-Länge einer Bibliothek ( $AIS$ ) betrachtet wurde. Daraufhin wurde in Anlehnung an die Illumina Guidelines der UCI Genomics High Throughput Facility [354] Formel 1 dazu verwendet die Molarität in nM ( $nM$ ) jeder fertigen Sequenzierbibliothek zu berechnen.

$$nM = \frac{conc}{(AIS + 120) * 650 \frac{g}{mol * bp}} * 10^6 \quad (1)$$

Im Anschluss wurde die Molarität ( $nM$ ) jeder fertigen Sequenzierbibliothek mit dem Faktor 1,3 multipliziert, da aufgrund vorhergehender Genomsequenzierungen und dem Remapping von Read-Daten (Methoden-Abschnitt 7. ) erkannt worden war, dass die Bioanalyzer Plattform Insert-Längenverteilungen von Sequenzierbibliotheken fehlerhaft wiedergibt.

Im späteren Verlauf stellte sich heraus, dass diese dem Bioanalyzer System inhärente fehlerhafte Bestimmung durch einfache lineare Regression modelliert werden kann, um die durchschnittliche Insert-Länge einer Sequenzierbibliothek präziser abschätzen zu können. Der lineare Zusammenhang des Modells ist in Formel 2 dargestellt und sollte in Verbindung mit Formel 1 genutzt werden, um die präzisesten Ergebnisse bei der Bestimmung der durchschnittlichen Insert-Länge einer Sequenzierbibliothek zu erhalten.

$$AIS = 0,564 * AIS_{Bio} + 258 \quad (2)$$

### 6.1.8. Pooling und Laden von Sequenzierbibliotheken

Die gleichzeitige (parallele) Sequenzierung mehrerer Bibliotheken in einem Sequenzierlauf ermöglicht es ein gutes Verhältnis zwischen den anfallenden Kosten und der generierten Datenmenge pro sequenziertem Genom zu erzielen. Hierbei müssen zunächst mehrere unterschiedliche Bibliotheken in einem Pool zusammengefasst werden. Dabei ist darauf zu achten, dass die Adapter der Bibliotheken ein und desselben Pools jeweils unterschiedliche Indizes aufweisen. Die Indizierung ermöglicht die richtige Zuordnung (Demultiplexing) der sequenzierten DNA Fragmente (Read-Daten) zu den Bibliotheken (Genomen) aus denen sie ursprünglich stammten.

Pro Sequenzierlauf wurde eine Auswahl fertiger Sequenzierbibliotheken auf 1 nM verdünnt und zu einem Pool zusammengeführt. Dem Pool wurde das gleiche Volumen (Verhältnis 1:1) von frisch vorbereitetem 0,1 N NaOH hinzugefügt. Danach wurde der Pool durch vorsichtiges Auf- und Abpipettieren gemischt, bevor er auf einem vorgeheizten QBT 2 Heizblock für 5 min bei 95 °C inkubiert wurde. Die Behandlung des Pools mit Natronlauge und Hitze führt zur Denaturierung der doppelsträngigen DNA Fragmente zu Einzelsträngen. Im Anschluss wurde der Pool sofort auf Eis gelegt und das gleiche Volumen von frisch vorbereitetem und gekühltem 0,1 N HCl (Verhältnis 1:1:1) hinzugefügt. Der kühl gehaltene Pool wurde daraufhin durch vorsichtiges Auf- und Abpipettieren gemischt. Das rasche Abkühlen des Pools verhindert die Renaturierung der einzelsträngigen DNA Fragmente zu Doppelsträngen. Das Hinzufügen von Salzsäure hat eine Reduktion der Konzentration an Natronlauge aufgrund der Bildung von Natriumchlorid zur Folge. Dies ist notwendig, damit die spätere Hybridisierung der einzelsträngigen DNA Fragmente an den Oligonukleotid-Rasen der Flusszelle (flow cell) nicht unterbunden wird. Abschließend wurde der Pool mit vorgekühltem Hybridization Buffer (Illumina) auf die gewünschte Molarität (pMol) für die Sequenzierung verdünnt.

Pools bestehend aus IS1-Bibliotheken wurden bei 30-35 pMol sequenziert. Zusammengefasste IS2-Bibliotheken wurden bei 60-80 pMol sequenziert. IS3 Pools bei 150 pMol. Quantitative PCR (qPCR) wurde nicht ausgeführt, da die Berechnung der Molarität der einzelnen Bibliotheken eines Pools vor dem Poolen unter Verwendung der Formeln 1 und 2 ermittelt wurde. Damit entspricht die auf die

Flusszelle geladene DNA Menge der Menge an gesamter DNA im Pool und nicht nur der Menge der DNA Fragmente, die im Pool doppelt ligiert vorliegen, wie es bei der Konzentrationsermittlung via qPCR der Fall wäre.

## **6.2. Genomsequenzierungen für die Optimierung der PCR-freien Probenvorbereitung**

Im Rahmen der Optimierung der PCR-freien Probenvorbereitung wurden insgesamt 17 Bibliotheken hergestellt und sequenziert. Diese sind in Tabelle 3 zusammengefasst. Die Sequenzierungen fanden ausschließlich auf der MiSeq Sequenzierplattform (Illumina) statt. Von den 17 Bibliotheken wurden 14 mit einer Read-Länge von 2x200 bp (paired-end) sequenziert, wobei entweder v2 (500-cycle) oder v3 (600-cycle) MiSeq Reagent Kits (Illumina) verwendet wurden. Die restlichen 3 Bibliotheken (Mlu<sub>50</sub>\_IS2, Mlu<sub>50</sub>\_IS3 und Mlu<sub>50</sub>\_IS4) wurden mit einer Read-Länge von 2x25 bp unter Verwendung eines v2 (50-cycle) MiSeq Reagent Kits sequenziert. Das Demultiplexing der Read-Daten sequenzierter Pools wurde mit Hilfe der on-board MiSeq Reporter Software (v.2.3.32) ausgeführt. Die Rohdaten der Sequenzierungen wurden im Sequencing Read Archive (SRA, <http://www.ncbi.nlm.nih.gov/sra>) unter den Kennungen SRX1160338 (Bce\_TS), SRX1161452 (Bce\_IS1), SRX1161508 (Bce\_IS3), SRX1161510 (Efa\_TS), SRX1161512 (Efa\_IS1), SRX1161513 (Efa\_IS2), SRX1161514 (Sen\_IS1), SRX1161515 (Sen\_IS2), SRX1161540 (Pst\_TS), SRX1161638 (Pst\_IS1), SRX1161639 (Pst\_IS3), SRX1161640 (Mlu\_TS), SRX1161641 (Mlu\_IS1), SRX1161642 (Mlu\_IS3), SRX1161643 (Mlu<sub>50</sub>\_IS2), SRX1161644 (Mlu<sub>50</sub>\_IS3) und SRX1161645 (Mlu<sub>50</sub>\_IS4) hinterlegt. Auf sämtliche Rohdaten kann auch über das NCBI Bioproject PRJNA293409 (<http://www.ncbi.nlm.nih.gov/bioproject>) zugegriffen werden.

## **6.3. Genomsequenzierungen für die Speziesabgrenzung**

Tabelle 2 umfasst alle bakteriellen Stämme, die im Rahmen der Speziesabgrenzung innerhalb der Gattung *Pseudomonas* und der *B. cereus s.l.* Gruppe selbst sequenziert wurden. Für jeden dieser Stämme wurde mindestens eine Bibliothek der Kategorie IS1 hergestellt (Methoden-Abschnitt 6.1. ). Die Sequenzierung (paired-end) der Bibliotheken fand auf der MiSeq Sequenzierplattform (Illumina) statt. Tabelle 15 (Seite 80) und Tabelle 20 (Seite 107) gibt Aufschluss darüber wie viele Bibliotheken pro Stamm hergestellt und mit welcher Read-Länge sequenziert wurden. Für die Sequenzierungen wurden vorrangig v3 (600-cycle) MiSeq Reagent

Kits (Illumina) verwendet. Die Bibliotheken der Stämme *P. brenneri* DSM 15294<sup>T</sup>, *P. gessardii* DSM 17152<sup>T</sup>, *P. panacis* DSM 18529<sup>T</sup>, *P. panacis* WS 4668, *P. proteolytica* DSM 15321<sup>T</sup>, *P. rhodesiae* DSM 14020<sup>T</sup>, *P. rhodesiae* WS 4669, *P. veronii* DSM 11331<sup>T</sup>, *P. veronii* DSM 16272, *P. veronii* WS 4670 und *P. veronii* WS 4671 wurden dahingegen mit v2 (500-cycle) MiSeq Reagent Kits (Illumina) sequenziert. Das Demultiplexing nach der Sequenzierung von Bibliotheken-Pools wurde durch die on-board MiSeq Reporter Software (v2.4.60.8) durchgeführt.

#### 6.4. Qualitätskontrolle der Rohdaten der Genomsequenzierungen

Das Trimmen der Rohdaten erfolgte mit dem NGS QC Toolkit (v2.2.3) [355]. Dabei wurde das Skript TrimmingReads.pl verwendet. Bibliotheken, die im Rahmen der Optimierung der PCR-freien Probenvorbereitung hergestellt und mit einer Read-Länge von 2x200 bp sequenziert wurden (Tabelle 3), wurden immer um 10 Nukleotide vom 5'-Read-Ende und 1 Nukleotid vom 3'-Read-Ende aus getrimmt. Die sich daraus ergebenden (ungefilterten) Bibliotheken enthielten somit immer 2x189 bp paired-end Reads. Die drei Bibliotheken (Mlu<sub>50</sub>\_IS2, Mlu<sub>50</sub>\_IS3 und Mlu<sub>50</sub>\_IS4), welche mit einer Read-Länge von 2x25 bp sequenziert wurden, wurden dahingegen nicht getrimmt. Die Genomsequenzierungen innerhalb der Gattung *Pseudomonas* (Tabelle 15) und der *B. cereus s.l.* Gruppe (Tabelle 20) wurden teils mit unterschiedlichen Read-Längen durchgeführt. Alle Bibliotheken wurden nach ihrer Sequenzierung einheitlich 10 Nukleotide vom 5'-Read-Ende und mindestens 1 Nukleotid vom 3'-Read-Ende aus getrimmt. Das Trimmen der 3'-Enden war somit variabel und abhängig von der erzielten Gesamtqualität des jeweiligen Sequenzierlaufes.

Neben dem Trimmen der Rohdaten wurde das NGS QC Toolkit auch für das Filtern hoch-qualitativer Read-Paare aus den Rohdaten verwendet. Dies fand mit dem Skript IlluQC\_PRL.pl statt, wobei die automatische Detektion des Formates der FASTQ-Eingabedatei (Option 'A') und die Multiplexing DNA Library (Option '6') zum Entfernen Adapter-kontaminierter Reads verwendet wurde. Generell wurden die Rohdaten zuerst getrimmt, bevor nach hoch-qualitativen Read-Paaren gefiltert wurde. Sofern nicht anders angegeben, bestand ein einzelner Read die Filter-Prozedur, wenn mindestens 80 % seiner Länge (Parameter '-l 80.0') einen Phred Qualitätswert  $\geq 20$  (Parameter '-s 20') aufwies (80;20). Reads, die durch die Filter-Prozedur ihr vorwärts- (forward Read) oder rückwärts-gerichtetes (reverse Read) Gegenstück verloren hatten, wurden ebenso nicht mehr weiter beachtet und

verworfen. Das Tool FastQC (v0.10.1, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) wurde für die Kontrolle und Visualisierung der Qualität der Rohdaten, der getrimmten Rohdaten und der gefilterten hoch-qualitativen Read-Paare herangezogen.

### **6.5. De novo Genomassemblierung und Validierung der Assemblierungen**

Die im Rahmen der Optimierung der PCR-freien Probenvorbereitung durchgeführten Assemblierungen und deren Validierungen waren Teil einer größeren Strategie zur Evaluierung von Faktoren, die die Assemblierung bakterieller Genome beeinflussen (Methoden-Abschnitt 8. ). Die im Zuge dieser Strategie durchgeführten Assemblierungen und Validierungen werden in Methoden-Abschnitt 8.3. und Methoden-Abschnitt 8.4. genauer erläutert. Im Folgenden wird lediglich auf die Assemblierung der Genome, welche im Rahmen der Speziesabgrenzung innerhalb der Gattung *Pseudomonas* (Tabelle 15) und der *B. cereus s.l.* Gruppe (Tabelle 20) sequenziert wurden, eingegangen.

#### **6.5.1. De novo Genomassemblierung (Speziesabgrenzung)**

Die Assemblierung der Genome aus Tabelle 15 (*Pseudomonas*) und Tabelle 20 (*B. cereus s.l.*) fand nach der Qualitätskontrolle und somit ausgehend von gefilterten Bibliotheken (mit hoch-qualitativen Read-Paaren) statt. Für die Assemblierung wurde ausschließlich der Assembler SPAdes (v2.5.1) [356] verwendet, wobei die read error correction Funktionalität (BayesHammer [357]) und die contig mismatch correction Funktionalität (Parameter '--careful') des Assemblers genutzt wurde. Die k-mer Kombination (Parameter '-k') wurde unabhängig von der Read-Länge der jeweiligen Bibliothek auf '21, 33, 55, 77, 99, 127' gesetzt. Bei Genomen für die zwei Sequenzierbibliotheken hergestellt wurden, waren beide Bibliotheken Teil der jeweiligen Genomassemblierung. Generell wurden Contigs mit einer Länge < 500 bp aus den Assemblierungen entfernt. Zudem wurden die Assemblierungen hin auf eine Kontamination mit Sequenzen aus dem Bakteriophagen  $\phi$ X174 (RefSeq NC\_001422.1) mittels Standard Nucleotide BLAST (blastn) [49] überprüft. Bei Kontamination wurden die Phagen-Sequenzen aus den Assemblierungen entfernt, was gegebenenfalls das Aufbrechen von Contigs nach sich zog. Abschließend wurden die Genomassemblierungen unter Anwendung bestimmter Assemblierungs-Metriken validiert.

### 6.5.2. Validierung der Genomassemblierungen (Speziesabgrenzung)

Für die Validierung der in Tabelle 15 (*Pseudomonas*) und Tabelle 20 (*B. cereus s.l.*) aufgeführten Genomassemblierungen wurden folgende Metriken herangezogen: Sequenziertiefe, Längstes Contig, N50, L50, Gesamtanzahl der Contigs und die Assembly Size. Die Sequenziertiefe  $C$  wurde definiert als  $C = \sum_{i=1}^n \left( \left\lceil \frac{2L_i N_i}{G} \right\rceil \right)$ . Dabei ist  $n$  die Anzahl der verwendeten Bibliotheken für die Genomassemblierung.  $L_i$  entspricht der Read-Länge der  $i$ -ten Bibliothek in bp und  $N_i$  entspricht der Anzahl der hoch-qualitativen Read-Paare der  $i$ -ten Bibliothek.  $G$  ist die Assembly Size, die der Gesamtlänge aller Contigs der Genomassemblierung in bp entspricht. N50 entspricht der Länge des Contigs, das 50 % der Assembly Size überschreitet, nachdem die Menge aller Contigs vom längsten zum kürzesten hin sortiert und summiert wurde [358]. L50 ist die Anzahl an Contigs, deren summierte Länge 50 % der Assembly Size überschreitet, nachdem die Menge aller Contigs vom längsten zum kürzesten hin sortiert wurde.

## 7. Remapping ungefilterter und gefilterter Sequenzierbibliotheken

Die im Rahmen der Optimierung der PCR-freien Probenvorbereitung sequenzierten Bibliotheken der Kategorien IS1-IS3 (Tabelle 3) wurden zur Bestimmung ihrer Insert-Längenverteilungen vor (ungefilterte Bibliotheken) und nach der Qualitätskontrolle (gefilterte Bibliotheken) gegen ihre jeweiligen Referenzgenomsequenzen (Tabelle 1) gemappt. Dazu wurde das Burrows-Wheeler Alignment Tool (v0.6.2) [359] verwendet. Die Konstruktion der FM-Indices der Referenzgenome (Befehl 'index') sowie das Alignment der Read-Paare (Befehle 'aln' und 'sampe') fanden unter Verwendung der Standardparametereinstellungen statt. Einzige Ausnahme war die Begrenzung der Anzahl zurückgegebener Alignments für korrekt und nicht-übereinstimmend gepaarte Reads, die auf 1 (Parameter '-n 1' und '-N 1' des Befehles 'sampe') gesetzt wurde. Zudem wurde der gesamte FM-Index des jeweiligen Referenzgenomes in den Arbeitsspeicher geladen, um unnötige Festplattenzugriffe zu vermeiden (Parameter '-P' des Befehles 'sampe'). Die Read-Alignments wurden als Ergebnisdateien im SAM (Sequence Alignment/Map) Format ausgegeben. Die SAM-Dateien wurden anschließend mit dem Skript CleanSam bereinigt, bevor sie mit dem Skript SortSam nach Koordinaten (Parameter 'SORT\_ORDER=coordinate') sortiert wurden. Abschließend wurden die Insert-Längenverteilungen der ungefilterten und gefilterten Bibliotheken basierend auf den

entsprechenden SAM-Dateien (nun bereinigt und nach Koordinaten sortiert) mit dem Skript `CollectInsertSizeMetrics` berechnet. Die Skripte `CleanSam`, `SortSam` und `CollectInsertSizeMetrics` wurden generell mit milder Validierungsstrengung (Parameter `'VALIDATION_STRINGENCY=LENIENT'`) aufgerufen und stammten aus dem Picard Toolkit (v1.84) [360].

## **8. Evaluierung von Faktoren, die die *de novo* Assemblierung bakterieller Genome beeinflussen**

### **8.1. Analysierte Faktoren**

Der Einfluss des Assemblers, der Insert-Länge, der Read-Länge und der Sequenziertiefe (sequencing depth) auf die Assemblierung bakterieller Genome wurde untersucht. Die Grundlage der Evaluierung bildeten die 14 ungefilterten Bibliotheken, die im Rahmen der Optimierung der PCR-freien Probenvorbereitung mit einer Read-Länge von 2x200 bp sequenziert (Tabelle 3) und anschließend auf 2x189 bp getrimmt wurden (Methoden-Abschnitt 6.4. ).

Die getesteten Assembler waren ABySS (v1.3.3) [361], Edena (v3.131028) [362], SPAdes (v2.5.1) [356] und Velvet (v1.2.10) [363]. Zu den untersuchten Kategorien (durchschnittlichen Insert-Längen) von Sequenzierbibliotheken zählten TS (Bce, Efa, Pst), IS1 (Bce, Efa, Sen, Pst), IS2 (Efa, Sen) und IS3 (Bce, Pst). Die überprüften Read-Längen betragen 100, 125, 150, 175 und 189 Nukleotide. Die Anpassung der Read-Länge wurde durch wiederholtes Trimmen der 3'-Enden der ungefilterten Bibliotheken mit einer Read-Länge von 2x189 bp bewerkstelligt. Die betrachteten Sequenziertiefen waren 45- und 90-fach. Teil-Bibliotheken mit unterschiedlichen Sequenziertiefen wurden durch randomisiertes Ziehen von Read-Paaren aus den auf die jeweilige Read-Länge getrimmten und anschließend gefilterten Bibliotheken erzeugt. Das Filtern der Bibliotheken fand wie in Methoden-Abschnitt 6.4. mit den Grenzwerten 80;20 statt.

### **8.2. Erstellung von Teil-Bibliotheken**

Aus jeder ungefilterten Bibliothek mit einer Read-Länge von 2x189 bp wurden mehrere gefilterte Teil-Bibliotheken mit unterschiedlichen Read-Längen und Sequenziertiefen erzeugt (Methoden-Abschnitt 8.1. ). Für jede Read-Länge und Sequenziertiefe wurden jeweils 3 unabhängige Teil-Bibliotheken erstellt. Dies resultierte in einer Gesamtmenge von 300 Teil-Bibliotheken bestehend aus 100

Triplikaten für die Genome Bce, Efa, Sen und Pst. Die Bibliotheken des Stammes Mlu wurden nicht zur Erstellung derartiger Teil-Bibliotheken herangezogen. Somit ergibt sich für die Genome Bce, Efa und Pst jeweils eine Anzahl von  $3 * 5 * 2 * 3 = 90$  generierten Teil-Bibliotheken. Dies ist darauf zurückzuführen, dass 3 Kategorien, 5 Read-Längen und 2 Sequenziertiefen in drei voneinander unabhängigen Ausfertigungen (Triplikate) für jedes dieser Genome betrachtet wurden. Für Sen ergab sich eine Anzahl von  $2 * 5 * 1 * 3 = 30$  erzeugten Teil-Bibliotheken, da lediglich zwei Bibliotheken unterschiedlicher Kategorie (IS1, IS2) für dieses Genom hergestellt und nur eine Sequenziertiefe (45-fach) untersucht wurde.

### **8.3. De novo Genomassemblierung der Teil-Bibliotheken**

Jede der 300 Teil-Bibliotheken (Methoden-Abschnitt 8.2. ) wurde mit den vier Assemblern ABySS (v1.3.3), Edena (v3.131028), SPAdes (v2.5.1) und Velvet (v1.2.10) assembliert. Somit wurden insgesamt  $4 * 300 = 1.200$  (optimale) Genomassemblierungen durchgeführt. Sofern nicht anders angegeben, fanden alle Assemblierungen unter den Standardparametereinstellungen des jeweiligen Assemblers statt. Im Allgemeinen wurden Contigs mit einer Länge  $< 500$  bp aus den Assemblierungen entfernt.

#### **8.3.1. Genomassemblierungen mit ABySS**

Für jede Teil-Bibliothek wurden mehrere ABySS Assemblierungen in einem Intervall von k-mer Werten durchgeführt. Dabei wurde der k-mer Wert schrittweise um 2 erhöht. Die untere Grenze des k-mer Intervalls wurde immer auf 21 gesetzt. Die obere Grenze des k-mer Intervalls war abhängig von der Read-Länge der zu assemblierenden Teil-Bibliothek. Für Bibliotheken mit einer Read-Länge von 189, 175, 150, 125 und 100 Nukleotiden wurde die obere k-mer Grenze entsprechend auf 171, 161, 141, 111 und 91 gesetzt. Für jede Teil-Bibliothek wurde lediglich die optimale Genomassemblierung weiter berücksichtigt und alle anderen verworfen. Die optimale Genomassemblierung galt als diejenige mit dem höchsten N50 Wert. Sofern mehrere k-mer Werte zu Assemblierungen mit maximalem N50 Wert führten, wurde diejenige mit der geringsten Anzahl an Contigs als optimal angesehen. Gab es auch hier Gleichstände wurde die Assemblierung mit der längsten durchschnittlichen Contig-Länge als optimal betrachtet.

### 8.3.2. Genomassemblierungen mit Edena

Für jede Teil-Bibliothek wurde ein overlaps graph mit einer minimalen Überlappungslänge von 21 Nukleotiden vorberechnet (Parameter '-M 21'). Ein Kürzen der 3'-Enden der Reads auf eine bestimmte Read-Länge fand dabei nicht statt (Parameter '-t' nicht angewendet). Auf der Grundlage des vorberechneten overlaps graph wurden für jede Teil-Bibliothek mehrere Edena Assemblierungen in einem Intervall von overlap Werten (Parameter '-m') berechnet. Der overlap cutoff wurde dabei schrittweise um 2 erhöht. Der minimale overlap cutoff entsprach immer 21 Nukleotiden. Die maximale Suchdistanz für Read-Paare wurde zudem immer auf 5.000 bp begrenzt (Parameter '-peHorizon 5000'). Die obere Grenze des overlap cutoffs Intervalls war abhängig von der Read-Länge der jeweiligen Teil-Bibliothek. Für Read-Längen von 189, 175, 150, 125 und 100 Nukleotiden wurden die maximalen overlap cutoffs entsprechend auf 161, 151, 131, 111 und 91 gesetzt. Für jede Teil-Bibliothek wurde nur die optimale Genomassemblierung (wie in Methoden-Abschnitt 8.3.1. definiert) weiter berücksichtigt und alle anderen Assemblierungen verworfen.

### 8.3.3. Genomassemblierungen mit SPAdes

Die Assemblierungen mit SPAdes wurden wie in Methoden-Abschnitt 6.5.1. durchgeführt. Allerdings war die angewendete k-mer Kombination (Parameter '-k') abhängig von der Read-Länge der jeweiligen Teil-Bibliothek. Bei Read-Längen von 189, 175 und 150 Nukleotiden wurde die k-mer Kombination auf '21, 33, 55, 77, 99, 127' gesetzt. Bei einer Read-Länge von 125 Nukleotiden war die k-mer Kombination '21, 33, 55, 77, 99'. Bei 100 Nukleotiden entsprach sie '21, 33, 55, 77'.

### 8.3.4. Genomassemblierungen mit Velvet

Die Assemblierung der Teil-Bibliotheken mit Velvet fand abhängig von der jeweiligen Read-Länge in den gleichen k-mer Intervallen wie in Methoden-Abschnitt 8.3.1. beschrieben statt. Auch die Definition und Auswahl der optimalen Genomassemblierung einer Teil-Bibliothek entsprach der aus Methoden-Abschnitt 8.3.1. Alle Velvet Assemblierungen wurden mit den Parametereinstellungen '-alignments yes', '-exp\_cov auto', '-ins\_length auto', '-scaffolding no' und '-read\_trkg yes' durchgeführt.

#### 8.4. Validierung der Genomassemblierungen

Häufig werden die Metriken N50 (Methoden-Abschnitt 6.5.2. ) oder NG50 für die Bewertung von Genomassemblierungen herangezogen [364]. Der NG50 Wert ist definiert wie der N50 Wert, wobei anstatt der Assembly Size die Gesamtlänge eines Referenzgenomes zur Berechnung herangezogen wird [98]. Unglücklicherweise sagt eine Validierung mit Hilfe des N50/NG50 Wertes eher etwas über die Längenverteilung der Contigs (Kontiguität) einer Genomassemblierung aus als über deren Genauigkeit/Verlässlichkeit [358]. In Fällen bei denen jedoch ein Referenzgenom vorhanden ist, kann die Menge der assemblierten Contigs gegen die Referenzsequenz aligniert werden. Dabei werden Contigs an ihren misassemblierten Regionen aufgebrochen, bevor der NG50 Wert berechnet wird. Diese Art der Vorgehensweise wurde bei den Genome Assembly Gold-Standard Evaluations (GAGE) in Form der corrected assembly contiguity metric eingeführt [94]. In dieser Arbeit wird diese Metrik als korrigierter NG50 Wert bezeichnet. Eine ähnliche Metrik mit der Bezeichnung NGA50 wird durch das Quality Assessment Tool for Genome Assemblies (QUAST) angewendet [365].

Die 300 Teil-Bibliotheken (bestehend aus 100 Triplikaten) wurden mit vier verschiedenen Assemblern (ABYSS, Edena, SPAdes, Velvet) assembliert, was zu einer Gesamtanzahl an 1.200 (optimalen) Genomassemblierungen führte. Diese 1.200 Assemblierungen wurden pro Assembler in jeweils 100 Assemblierungs-Sets aufgeteilt. Dabei enthielt jedes Assemblierungs-Set die 3 Genomassemblierungen eines bestimmten Teil-Bibliothek Triplikates mit einem bestimmten Assembler. Daraufhin wurden alle Assemblierungs-Sets des gleichen Assemblers gegen ihre Referenzgenomsequenzen aus Tabelle 1 mit QUAST (v2.2) [365] validiert. Dabei wurde für jede der 3 Genomassemblierungen eines Assemblierungs-Sets der korrigierte NG50 Wert sowie der NGA50 Wert abgespeichert. Anschließend wurde jedes Assemblierungs-Set zu einem Validierungseintrag reduziert, wobei jeder Validierungseintrag die beiden Durchschnittswerte der korrigierten N50 Werte sowie der NGA50 Werte seines Assemblierungs-Sets enthielt. So wurden insgesamt 400 Validierungseinträge (100 pro Assembler) erzeugt. Dabei war jeder Validierungseintrag eindeutig und einzigartig in Bezugnahme auf sein zugrundeliegendes Referenzgenom sowie der Kombination seiner Ausprägung der vier unterschiedlichen Faktoren (Assembler, Insert-Länge, Read-Länge und

Sequenziertiefe). Generell wurde die Validierung mit QUAST unter Anwendung der Parametereinstellungen ‘--ambiguity-usage one’ und ‘--gag’ durchgeführt.

### 8.5. Berechnung des relativen Assemblierungs-Wertes

Grundsätzlich wurden die Validierungs-Einträge nach den Ausprägungen des Faktors unter aktueller Beobachtung gruppiert. Dann wurde für jede Ausprägung der durchschnittliche korrigierte NG50 Wert und der durchschnittliche NGA50 Wert gebildet. Abschließend wurde der relative Assemblierungs-Wert  $S$  jeder Ausprägung für jede der beiden betrachteten Metriken (korrigierter NG50 und NGA50) bestimmt. Dies wurde dadurch bewerkstelligt, dass der Durchschnittswert jeder Ausprägung durch den Durchschnittswert der Ausprägung mit dem maximalen Durchschnittswert geteilt wurde. Somit ergibt sich, dass  $0 < S \leq 1$  ist. Je näher  $S$  dem Wert 1 kommt, desto höher ist der positive Effekt der entsprechenden Ausprägung des betrachteten Faktors auf die Genomassemblierung. Für die Evaluierung der Assembler wurden die Validierungs-Einträge in einer vom Referenzgenom abhängigen Weise betrachtet. Für alle anderen Faktoren (Insert-Länge, Read-Länge und Sequenziertiefe) wurden die Validierungs-Einträge pro Genom und Assembler analysiert.

## 9. Genomweite Vorhersage proteinkodierender Sequenzen

Prodigal (v2.60) [366] wurde für die Vorhersage der proteinkodierenden Sequenzen (CDS) aller offenen und geschlossenen Genomassemblierungen aus Tabelle 15, Tabelle 20, Tabelle 27, Tabelle 28, Tabelle 31 sowie der MLSA Außengruppen *Cellvibrio japonicus* Ueda107<sup>T</sup> (NC\_010995.1) und *Bacillus subtilis* ATCC 6051<sup>T</sup> (NC\_020507.1) im Normal-Modus (Parameter ‘-p single’) verwendet. Für die Vorhersage wurde grundsätzlich die Standard-Codon-Tabelle 11 (Parameter ‘-g 11’) des NCBI für Bakterien, Archaeen, prokaryotische Viren und Chloroplasten (<https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi>) benutzt. Zudem wurde die Vorhersage partieller CDS untersagt (Parameter ‘-c’) und das Scannen nach Motiven ribosomaler Bindestellen (Parameter ‘-n’) erzwungen. Für jede Genomassemblierung wurden die vorhergesagten CDS auf Nukleinsäure- (Parameter ‘-d’) und Aminosäureebene (Parameter ‘-a’) ausgegeben und für die weitere Analyse verwendet.

## 10. Multilocus-Sequenz-Analyse (MLSA)

### 10.1. Verwendete MLSA Schemata und Außengruppen

Im Falle der Gattung *Pseudomonas* wurde das MLSA Schema von Mulet *et al.* [25] verwendet. Dieses Schema umfasst die partiellen Nukleotidsequenzen der vier Gene 16S rDNA, *gyrB*, *rpoB* und *rpoD*. Die Sequenzbereiche der einzelnen Gene des MLSA Schemas sind nachfolgend in Tabelle 4 gelistet. Dabei sind die jeweiligen Anfangs- und Endnukleotide mit Bezug zum *P. aeruginosa* Typstamm angegeben. Im Gegensatz zum ursprünglichen Schema startet der 16S Sequenzbereich von Nukleotid 64, um eine Reduktion des *Pseudomonas* Datensatzes zu vermeiden. Somit ist das verwendete Schema 63 Nukleotide kürzer als das Original, das vom ersten Nukleotid aus startet. Als Außengruppe der Analyse wurde der Stamm *Cellvibrio japonicus* Ueda107<sup>T</sup> (NCBI Referenzgenomsequenz NC\_010995.1) gewählt.

Tabelle 4 Verwendetes MLSA Schema für die Gattung *Pseudomonas*

Gen	GenBank Sequenz ( <i>P. aeruginosa</i> <sup>T</sup> )	Startnukleotid (inklusive)	Endnukleotid (inklusive)	Sequenzlänge in Nukleotiden
16S	X06684	64	1.295	1.232
<i>gyrB</i>	AB039386	43	843	801
<i>rpoB</i>	AJ717442	40	954	915
<i>rpoD</i>	AB039607	43	759	717

Für die *B. cereus* s. l. Gruppe wurde ein MLSA Schema bestehend aus den sieben proteinkodierenden Genen *adk* (adenylate kinase), *ccpA* (catabolite control protein A), *glpF* (glycerol uptake facilitator), *glpT* (glycerol-3-phosphatetransporter), *panC* (pantoate- $\beta$ -alanine ligase), *pta* (phosphate acetyltransferase) und *pycA* (pyruvate carboxylase) ausgewählt [26, 367]. Obwohl das Schema auf der Verwendung der gesamten kodierenden Nukleotidsequenz aller sieben Gene beruht, sind aus Gründen der Übersicht und Vollständigkeit die Sequenzbereiche der einzelnen Gene mit Bezug zum *B. cereus* Typstamm in Tabelle 5 angegeben. Während dem Screening nach GlpT (Methoden-Abschnitt 10.2. ) war aufgefallen, dass das entsprechende Protein in elf Stämmen trunkiert oder gar nicht vorhanden war. Um Datenreduktion zu vermeiden, wurde das Gen nachträglich aus dem MLSA Schema genommen. Somit besteht das in dieser Arbeit verwendete Schema aus sechs anstatt sieben Genen. *Bacillus subtilis* ATCC 6051<sup>T</sup> (NCBI Referenzgenomsequenz NC\_020507.1) wurde als Außengruppe der *B. cereus* s.l. Gruppe herangezogen.

Tabelle 5 Verwendetes MLSA Schema für die *B. cereus s. l.* Gruppe

Gen	RefSeq Sequenz ( <i>B. cereus</i> <sup>T</sup> )	Position in RefSeq Sequenz	Startnukleotid (inklusive)	Endnukleotid (inklusive)	Sequenzlänge in Nukleotiden
<i>adk</i>		137.017 - 137.667	1	651	651
<i>ccpA</i>		4.611.711 - 4.612.709 (rk)	1	999	999
<i>glpF</i>	NC_004722.1	1.015.014 - 1.015.835	1	822	822
<i>panC</i>		1.489.715 - 1.490.563	1	849	849
<i>pta</i>		5.303.532 - 5.304.503 (rk)	1	972	972
<i>pycA</i>		3.926.851 - 3.930.297 (rk)	1	3.447	3.447

rk, reverse komplementierte Sequenz

## 10.2. Screening nach MLSA Genen in Genomassemblierungen

Das Screening nach den proteinkodierenden Genen beider MLSA Schemata (Tabelle 4 und Tabelle 5) verlief sowohl für die Stämme der Gattung *Pseudomonas* als auch der *B. cereus s. l.* Gruppe inklusive beider Außengruppen nach dem gleichen Prinzip. Zunächst wurde für jedes proteinkodierende Gen der MLSA-Schemata eine Query-Referenzsequenz auf Aminosäureebene vom NCBI bezogen (Tabelle 6). Danach wurde für jeden zu untersuchenden Stamm eine Protein-Referenzdatenbank der genomweit vorhergesagten CDS des Stammes (Methoden-Abschnitt 9. ) mit Hilfe des BLAST-Programmes makeblastdb (v2.2.25+) [368] angelegt. Mittels Protein-Protein BLAST (blastp v2.2.25+) [368] wurde schließlich der beste Treffer der jeweiligen Query-Referenzsequenz zur Protein-Referenzdatenbank des untersuchten Stammes ermittelt (Parameter '-num\_descriptions 1' und '-num\_alignments 1'). Die Ergebnisse aller einseitigen Bester-Treffer Screenings gegen alle Protein-Referenzdatenbanken (Stämme) wurden nochmals manuell auf ihre Richtigkeit hin überprüft, bevor die zugehörigen CDS auf Nukleinsäureebene (!) weitere Verwendung fanden. Nach dem Screening mussten die Stämme *P. stutzeri* YC-YH1, *B. cereus* H3081.97, *B. cereus* K-5975c und *B. mycoides* Rock1-4 aus allen weiteren Analysen ausgeschlossen werden. Der jeweilige Grund ihres Ausschlusses kann Tabelle 29 bzw. Tabelle 32 entnommen werden.

Tabelle 6 Verwendete Query-Referenzsequenzen (MLSA Screening)

Protein	RefSeq/GenBank Query-Referenzsequenz	Referenzstamm der Query-Referenzsequenz	MLSA Schema
GyrB	BAB17356.1	<i>P. aeruginosa</i> IFO 12689 (T)	Gattung <i>Pseudomonas</i> (Tabelle 4)
RpoB	CAG30614.1	<i>P. aeruginosa</i> LMG 1242 (T)	
RpoD	BAB17577.1	<i>P. aeruginosa</i> IFO 12689 (T)	
Adk	NP_830031.1	<i>Bacillus cereus</i> ATCC 14579 (T)	<i>B. cereus</i> s. l. Gruppe (Tabelle 5)
CcpA	NP_834378.1	<i>Bacillus cereus</i> ATCC 14579 (T)	
GlpF	NP_830820.1	<i>Bacillus cereus</i> ATCC 14579 (T)	
GlpT	NP_830470.1	<i>Bacillus cereus</i> ATCC 14579 (T)	
PanC	NP_831320.1	<i>Bacillus cereus</i> ATCC 14579 (T)	
Pta	NP_835048.1	<i>Bacillus cereus</i> ATCC 14579 (T)	
PycA	NP_833666.1	<i>Bacillus cereus</i> ATCC 14579 (T)	

Im Gegensatz zu den proteinkodierenden Genen (*gyrB*, *rpoB* und *rpoD*) des MLSA Schemas der Gattung *Pseudomonas* wurden mit Hilfe des Tools RNAmmer (v1.2) die bakteriellen 16S Sequenzen (Parameter '-S bac' und '-m ssu') direkt aus den Genomsequenzierungen vorhergesagt. Hierbei konnte es vorkommen, dass für einen *Pseudomonas* Stamm mehrere unterschiedliche 16S Sequenzen gefunden wurden. In diesem Fall wurde die am häufigsten gefundene 16S Sequenz verwendet. Sofern mehrere unterschiedliche 16S Sequenzen gleichhäufig vorkamen, wurde diejenige mit dem höchsten RNAmmer-Score gewählt. Für insgesamt 52 Stämme lieferte die Vorhersage keine oder lediglich zu kurze 16S Sequenzen. Ersatzweise konnte bei 32 Stämmen eine Sequenz über das NCBI bezogen werden (Tabelle 30). Alle restlichen Stämme wurden gänzlich von weiteren Analysen ausgeschlossen (Tabelle 29).

### 10.3. Vorbereitung und Alignment der verketteten MLSA Sequenzen

Insgesamt waren 460 *Pseudomonas* Stämme (P001-P460 in Tabelle 15, Tabelle 28 und Tabelle 30) zuzüglich der *C. japonicus* Ueda107<sup>T</sup> Außengruppe Teil der weitergehenden Analyse. Für jedes der vier MLSA Gene (16S rDNA, *gyrB*, *rpoB* und *rpoD*) wurde ein multiples Nukleotidsequenzalignment der zugehörigen MLSA Einzelgensequenzen (der Stämme) sowie der entsprechenden partiellen (!) *P. aeruginosa* Typstamm MLSA Referenzsequenz (Tabelle 4) mit dem Clustal Omega (v1.2.4) [369-371] Webservice (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) erstellt. Schließlich fand das Trimmen der Einzelgensequenzen auf den im Alignment zur partiellen *P. aeruginosa* Referenzsequenz getroffenen Sequenzbereich statt. Abschließend wurden die so erhaltenen partiellen Einzelgensequenzen für jeden Stamm in der Reihenfolge 16S rDNA, *gyrB*, *rpoB* und *rpoD* miteinander verkettet, bevor das endgültige multiple MLSA Sequenzalignment der verketteten Sequenzen

mit der ClustalW Funktionalität (DNA-Alignment mit Standardparametereinstellungen) [372] der MEGA 7 Software (v7.0.26) [373] errechnet wurde.

Neben der *B. subtilis* ATCC 6051<sup>T</sup> Außengruppe konnten insgesamt 281 Stämme (B001-B281 in Tabelle 20 und Tabelle 31) der *B. cereus s.l.* Gruppe für die weitere Analyse herangezogen werden. Da das MLSA Schema (Tabelle 5) auf der Verwendung der vollständigen CDS aller MLSA Einzelgensequenzen beruht, war ein Trimmen der Sequenzdaten nicht nötig. Für jeden untersuchten Stamm wurden die CDS (inklusive Stoppcodon) in der Reihenfolge *adk*, *ccpA*, *glpF*, *panC*, *pta* und *pycA* miteinander verkettet. Das finale multiple MLSA Sequenzalignment wurde wie bei den Stämmen der Gattung *Pseudomonas* erzeugt.

## 10.4. Phylogenetische Analyse

### 10.4.1. Auswahl des Substitutionsmodells

Mit Hilfe des Programmes jModelTest 2 (v2.1.10) [374, 375] wurde das geeignetste Nukleotid-Substitutionsmodell für das multiple MLSA Sequenzalignment der Gattung *Pseudomonas* und das der *B. cereus s.l.* Gruppe geschätzt. Dabei wurde unter Anwendung der Standardparametereinstellungen aus 88 Modellen mit und ohne Berücksichtigung ungleicher Basenhäufigkeiten (+F), Ratenvariation zwischen den Alignmentpositionen (+G, vier diskrete Gamma-Kategorien) und Anteilen nicht-variabler Alignmentpositionen (+I) das am besten angepasste Modell (best-fit model) durch das Akaike Informationskriterium (AIC) [376], das korrigierte Akaike Informationskriterium (AIC<sub>c</sub>) [377, 378] sowie das Bayesianische Informationskriterium (BIC) [379] ermittelt. Für beide multiple MLSA Sequenzalignments resultierte das General Time Reversible Model (GTR) [380] unter Annahme von Ratenheterogenität entlang der Alignmentpositionen (GTR+G+I) als das am besten geeignete Modell. Da sowohl im Falle der Gattung *Pseudomonas* ( $w_{GTR+G+I}^{AIC} = 1$ ,  $w_{GTR+G+I}^{AICc} = 1$ ,  $w_{GTR+G+I}^{BIC} = 1$ ) als auch bei der *B. cereus s.l.* Gruppe ( $w_{GTR+G+I}^{AIC} = 1$ ,  $w_{GTR+G+I}^{AICc} = 1$ ,  $w_{GTR+G+I}^{BIC} = 0,9928$ ) die AIC, AIC<sub>c</sub> und BIC Gewichtungen (weights)  $w$  Werte > 0,99 aufwiesen und somit die 95%-Modellkonfidenzmenge in beiden Fällen ausschließlich aus dem GTR+G+I Modell bestand, wurde auf eine weitergehende Modellmittelung (model averaging) verzichtet.

### 10.4.2. Phylogenetische Rekonstruktion und Visualisierung der MLSA Stammbäume

Auf Grundlage der multiplen Sequenzalignments der Gattung *Pseudomonas* und der *B. cereus s.l.* Gruppe fand die Rekonstruktion beider MLSA Stammbäume mit Hilfe der MEGA 7 Software (v7.0.26) [373] statt. Dabei wurde als statistische Methode die Maximum-Likelihood (ML) Schätzung gewählt. Die Prüfung der statistischen Verlässlichkeit der Knoten in beiden ML-Bäumen fand durch ein Bootstrap-Verfahren mit 500 Replikaten statt. Als Nukleotid-Substitutionsmodell wurde jeweils das GTR+G+I Modell (fünf diskrete Gamma-Kategorien) bestimmt. Alle weiteren Parametereinstellungen wurden auf den vorgegebenen Standardeinstellungen belassen. Die Visualisierung der MLSA ML-Stammbäume erfolgte mit dem Online-Tool Interactive Tree Of Life (iTOL, v3.6.1) [381].

## 11. Paarweise Ähnlichkeitsmaße

### 11.1. Berechnung und Vergleich von MLSA Sequenzähnlichkeiten

Die prozentuale MLSA Sequenzähnlichkeit wurde ausschließlich für Paare von Stämmen aus der Gattung *Pseudomonas* berechnet, da ein Spezies-Grenzwert von 97 % bereits vorgeschlagen und bekräftigt wurde [25, 125]. Grundlage der Berechnung war das multiple Nukleotidalignment der verketteten partiellen Sequenzen der Gene 16S rDNA, *gyrB*, *rpoB* und *rpoD* (Methoden-Abschnitt 10.3. ). Das Skript dnadist des Phylogeny Inference Package (PHYLIP v3.696) wurde zur Ermittlung der paarweisen Sequenzähnlichkeiten genutzt.

### 11.2. Arten und Berechnung paarweiser genomischer Ähnlichkeitsmaße

Zur Bestimmung des evolutionären Verwandtschaftsgrades zweier Genome wurden die Ähnlichkeitsmaße ANI<sub>b</sub> (Average Nucleotide Identity mittels blast) [47, 48], ANI<sub>m</sub> (Average Nucleotide Identity mittels MUMmer) [48] sowie gANI (genome-wide Average Nucleotide Identity) [52] verwendet.

Grundsätzlich unterscheiden sich die Algorithmen zur Berechnung der paarweisen Genomähnlichkeiten in der Sequenzart (Genomsequenzen vs. proteinkodierende Sequenzen), die zum Vergleich herangezogen wird. Außerdem gibt es Unterschiede bei den Alignment-Algorithmen mit Hilfe derer der Vergleich stattfindet, was Auswirkungen auf die Schnelligkeit der Berechnung zur Folge hat [48, 52].

Gemeinsam haben alle Ähnlichkeitsmaße, dass bei ihrer Berechnung von zwei zu vergleichenden Genomen *A* und *B* ausgegangen wird. Dabei wird ein Genom als

Query-Genom und das andere als Referenz-Genom definiert. Die Ähnlichkeitsmaße verhalten sich in der Regel nicht symmetrisch. Beispielsweise gilt

$$ANib(A_{Query}, B_{Referenz}) \neq ANib(B_{Query}, A_{Referenz}).$$

### 11.2.1. ANIb (Average Nucleotide Identity mittels blast)

Der Algorithmus zur Berechnung des ANIb Wertes wurde von Goris *et al.* [47] entwickelt. Im ersten Schritt wird die Sequenz des Query-Genomes in konsekutive DNA-Fragmente mit einer Länge von 1.020 Nukleotiden gesplitted. Daraufhin werden die DNA-Fragmente des Query-Genomes mittels blastn [49] gegen das Referenz-Genom aligniert, wobei für jedes DNA-Fragment nur der beste Hit gegen das Referenz-Genom weiter betrachtet wird. Anschließend werden die (besten) Hits nach jenen gefiltert die eine Sequenzidentität  $\geq 30\%$  gerechnet auf die Gesamtlänge des DNA-Fragmentes aufweisen und das Alignment des Hits  $\geq 70\%$  der Länge des DNA-Fragmentes beinhaltet. Abschließend wird der ANIb Wert als die durchschnittliche Sequenzidentität über alle gefilterten Hits berechnet.

Die Berechnung des ANIb fand mit zwei unterschiedlichen Implementierungen statt. Zum einen wurde das Skript `calculate_ani.py` von Leighton Pritchard (<https://github.com/widdowquinn/scripts>) mit den BLAST-Programmen `makeblastdb` (v2.2.25+) und `blastn` (v2.2.25+) [368] verwendet. Zum anderen das Skript `ANI.pl` (Version vom 23.06.2013) von Jiapeng Chen (<https://github.com/chjp/ANI>) mit den BLAST-Programmen `formatdb` (v2.2.25) und `blastall` (v2.2.25) [49]. Die Implementierung von Leighton Pritchard ist als ANIb<sub>1</sub> gekennzeichnet. Die von Jiapeng Chen als ANIb<sub>2</sub>. Bei beiden Implementierungen verhält sich das Ergebnis der Berechnung des ANIb Wertes zweier Genome *A* und *B* nicht symmetrisch. Deshalb wurde der Durchschnitt der zwei gerichteten ANIb Werte als ANIb Wert des Genompaars *A, B* herangezogen. Somit gilt

$$ANib(\text{Genompaar } A, B) := \frac{ANib(A_{Query}, B_{Referenz}) + ANib(B_{Query}, A_{Referenz})}{2}.$$

### 11.2.2. ANIm (Average Nucleotide Identity mittels MUMmer)

Erstmalig wurde der Algorithmus zur Berechnung des ANIm Wertes von Richter *et al.* [48] vorgestellt. Dieser beruht auf der Verwendung des MUMmer Tools [50] und nutzt somit Suffixbäume zur Repräsentation von Genomsequenzen. Die Berechnung der paarweisen Alignments zwischen Query- und Referenz-Genom geschieht mit Hilfe von NUCmer [382]. Der prozentuale ANIm ergibt sich aus der Differenz der

Gesamtlänge aller NUCmer-produzierten Alignments mit der Gesamtanzahl der similarity errors (nicht-positive Match-Scores) in diesen Alignments. Die Berechnung des ANIm ist schneller als die des ANIb, da die Raum- und Zeitkomplexität für die Konstruktion und das Durchsuchen der Suffixbäume mit dem MUMmer System linear ist [383].

Für die Berechnung des ANIm wurde das Skript `calculate_ani.py` (<https://github.com/widdowquinn/scripts>) von Leighton Pritchard (MUMmer v3.23) herangezogen. Wie beim ANIb (Methoden-Abschnitt 11.2.1. ) wurde der Durchschnitt der zwei gerichteten ANIm Berechnungen als der ANIm Wert eines Genompaars  $A, B$  definiert.

### 11.2.3. AF, gANI (Alignment Fraction, genome-wide Average Nucleotide Identity)

Die Vorgehensweise zur Berechnung des AF und gANI Wertes eines Genompaars wurde von Varghese *et al.* [52] beschrieben. Ausgegangen wird dabei von den CDS auf Nukleinsäureebene des Query- und des Referenz-Genomes. Zu Beginn findet ein jeder-gegen-jeden Vergleich zwischen den CDS beider Genome mit NSimScan [384] statt. Dem folgt die Bestimmung aller Bidirektional-Besten-Hits (BBHs), die eine Sequenzidentität  $\geq 70\%$  sowie ein Alignment  $\geq 70\%$  (gemessen zur Proteinelänge des kürzeren Proteins) aufweisen. Der prozentuale gANI Wert des Query-Genomes mit dem Referenz-Genom ergibt sich schließlich aus der Gesamtanzahl der Identitäten aus den BBH-Alignments dividiert durch die Gesamtlänge der BBH CDS des Query-Genomes. Somit trifft der gANI Wert eine Aussage über den Ähnlichkeitsgrad der proteinkodierenden Information, die das Query-Genom mit dem Referenz-Genom teilt. Der AF Wert entspricht der Gesamtlänge der BBH CDS des Query-Genomes geteilt durch die Gesamtlänge aller CDS des Query-Genomes. Damit entspricht der AF Wert dem Anteil der zwischen Query- der proteinkodierenden Information und Referenzgenom geteilten proteinkodierenden Information zur Gesamtheit des Query-Genomes.

Alle paarweisen AF und gANI Werte wurden mit dem ANIcalculator (v1.0) der Microbial Species Identifier (MiSI) Software [52] unter Verwendung von NSimScan (v0.9.a) [384] berechnet. Sowohl die AF als auch der gANI verhalten sich nicht symmetrisch. Wie beim ANIb (Methoden-Abschnitt 11.2.1. ) ergibt sich in dieser Arbeit der gANI Wert eines Genompaars  $A, B$  aus dem Durchschnittswert der

beiden gerichteten gANI Berechnungen. Das gleiche gilt für den AF Wert, womit  $AF(\text{Genompaar } A, B) := \frac{AF(A_{\text{Query}}, B_{\text{Referenz}}) + AF(B_{\text{Query}}, A_{\text{Referenz}})}{2}$  ist.

## 12. Spezies-Grenzwert Ermittlung für die genomischen Ähnlichkeitsmaße ANIb und ANIm

In Analogie zum Vorgehen von Meier-Kolthoff *et al.* [59] wurde mit Hilfe binärer logistischer Regression [385] die Wahrscheinlichkeit, dass ein Genompaar eine DNA-DNA Hybridisierung (DDH)  $\geq 70$  % aufweist und somit zur gleichen prokaryotischen Spezies gehört statistisch modelliert. Ausgangspunkt waren die DDH Daten von 127 Genompaaren aus 27 prokaryotischen Gattungen. Davon stammten 62 aus Goris *et al.* [47], 36 aus Richter *et al.* [48] und 29 aus der Genomes OnLine Database (GOLD) [386]. Die Genompaare sowie dazugehörige DDH Werte sind in Tabelle 36 (Seite 218) aufgelistet. Für alle Genompaare bei denen eine experimentelle Bestimmung des DDH Wertes in beide Richtungen vorgenommen wurde, wurde der Durchschnitt aus den beiden DDH Werten herangezogen. Bei den restlichen Genompaaren wurde der jeweils vorhandene einseitige DDH Wert verwendet. Sofern DDH Werte  $\geq 100$  % auftraten, wurden diese auf 100 % gesetzt.

Für die genomischen Ähnlichkeitsmaße ANIb (beide Implementierungen) sowie ANIm wurde jeweils ein binomiales Logit-Modell geschätzt. Dabei nimmt die binäre abhängige Variable  $Y$  einen Wert von 1 ein, sofern ein Genompaar einen (durchschnittlichen) DDH Wert  $\geq 70$  % aufweist (ansonsten ist  $Y = 0$ ). Die unabhängige Variable  $X$  entspricht einem der untersuchten genomischen Ähnlichkeitsmaße (ANIb oder ANIm). Die 127 zur Modellbildung herangezogenen Ähnlichkeitswerte  $x_i$  jedes Ähnlichkeitsmaßes  $X$  wurden wie in Methoden-Abschnitt 11.2. beschrieben berechnet. Mit Hilfe der verschiedenen Logit-Modelle lässt sich für einen beliebigen paarweisen ANIb/ANIm Wert die Wahrscheinlichkeit, dass die beiden dem Vergleich zugrundeliegenden Stämme/Genome zur gleichen prokaryotischen Art gehören, schätzen.

Alle Logit-Modelle wurden in R [387] mit der Funktion `glm()` geschätzt (Tabelle 12 auf Seite 78). Mit dem Parameter "family=binomial" wurde dabei die Logit-Funktion als Linkfunktion spezifiziert. Die Signifikanz der Ähnlichkeitsmaße/Regressionskoeffizienten wurde in R mit dem Wald-Test (Funktion `summary()`) sowie dem Likelihood-Quotienten-Test (Funktion `pchisq()`) bestimmt

(Tabelle 13 auf Seite 78). Mit Hilfe der Pseudo- $R^2$ -Statistiken von McFadden [388], Cox & Snell [389] und Nagelkerke [390] konnte für jedes Logit-Modell die Gesamtmodellgüte ermittelt werden (Tabelle 14 auf Seite 78). Dafür wurde auf die Funktion LogRegR2() des R Packages descr (v1.1.3) [391] zurückgegriffen. Um die Klassifikationsgüte der Logit-Modelle einschätzen zu können, wurde die Area under curve (AUC) der entsprechenden Receiver-Operating-Characteristic-Kurven (ROC-Kurven) mit den Funktionalitäten aus dem R Package ROCR [392] berechnet (Tabelle 14).

### 13. Ermittlung und Visualisierung von Spezies-Clustern

Ausgangspunkt der Gruppierung einzelner Stämme in Spezies-Cluster waren die Ähnlichkeiten, die sich aus den paarweisen Vergleichen der Stämme ergaben. Grundsätzlich ergibt sich bei der Betrachtung von  $n$  Stämmen eine Anzahl von  $\frac{1}{2}(n^2 - n)$  zugrundeliegenden Vergleichen. Für die 460 betrachteten Stämme der Gattung *Pseudomonas* (P001-P460 in Tabelle 15, Tabelle 28 und Tabelle 30) ergaben sich somit 105.570 paarweise MLSA Sequenzähnlichkeiten. Da für 19 *Pseudomonas* Stämme keine Genomassemblierungen zur Verfügung standen (P442-P460 in Tabelle 30), belief sich die Anzahl der errechneten paarweisen Genomähnlichkeiten pro Index (ANIB<sub>11</sub>, ANIB<sub>12</sub>, ANIm, gANI) auf 97.020. Bei den 281 verwendeten Stämmen der *B. cereus s.l.* Gruppe (P001-P281 in Tabelle 20 und Tabelle 31) waren es entsprechend 39.340 paarweise Genomähnlichkeiten pro Index (ANIB<sub>11</sub>, ANIm, gANI). Für die Analyse dieser sehr großen Anzahl an Vergleichen wurde ein Graphen-basierter Ansatz gewählt, wobei Stämme als Knoten und paarweise Ähnlichkeiten als Kanten zwischen den Knoten in einem ungerichteten Graphen dargestellt wurden. Hierfür wurde mit Hilfe der Entwicklungsumgebung Eclipse Indigo (v3.7.2, <http://www.eclipse.org/>) ein Java-Programm (java-7-openjdk-amd64, <https://www.java.com/de/>) geschrieben, welches mit Hilfe der JDOM Bibliothek (v2.0.6, <http://www.jdom.org/>) aus der Gesamtinformation paarweiser Ähnlichkeiten jeweils eine Textdatei für die Gattung *Pseudomonas* und die *B. cereus s.l.* Gruppe im GraphML-Format (<http://graphml.graphdrawing.org/>) [393] erzeugt hat. GraphML ist ein auf der Extensible Markup Language (XML, <https://www.w3.org/XML/>) [394] basierendes Textdateiformat für die Darstellung und den Austausch von Graphen. Die beiden erzeugten GraphML-Dateien dienten als Eingabedateien der Graphen-Visualisierungs-Software Gephi (v0.9.1, <https://gephi.org/>) [395]. Unter Anwendung geeigneter Filter-Einstellungen in Gephi

wurde für jeden genomischen Index die Menge aller Kanten auf diejenigen paarweisen Ähnlichkeiten reduziert, die größer oder gleich dem anzuwendenden Spezies-Grenzwert (Tabelle 7) waren. Ähnlich zum Vorgehen von Varghese *et al.* [52] ergaben sich dadurch für jeden Index Spezies-Cluster in Form von Singletons, Cliques oder Cliques-Gruppen. Dabei entspricht eine Clique immer einer Menge von Knoten/Stämmen, bei der jedes Knotenpaar durch eine Kante miteinander verbunden ist (Ähnlichkeitswert jedes Stammpaares  $\geq$  Spezies-Grenzwert). Bei einer Cliques-Gruppe handelt es sich um eine Menge von zusammenhängenden Knoten/Stämmen, bei der jedoch nicht jeder Knoten mit jedem anderen Knoten verbunden ist. Zu beachten ist, dass sich eine Cliques-Gruppe somit immer aus mindestens zwei Cliques zusammensetzt. Knoten ohne Kanten werden als Singletons bezeichnet und entsprechen demnach Spezies-Clustern mit nur einem Mitglied/Stamm. Singletons können auch als einelementige Cliques angesehen werden. Mit Hilfe des Online-Tools Interactive Tree Of Life (iTOL, v3.6.1) [381] wurden die Spezies-Cluster, die sich für die Gattung *Pseudomonas* und die *B. cereus s.l.* Gruppe ergaben, in kodierter Form (Abbildung 11 auf Seite 95) zusammen mit dem jeweils zugehörigen MLSA ML-Stammbaum dargestellt.

**Tabelle 7 Angewendete Spezies-Grenzwerte zur Ermittlung von Spezies-Clustern**

Ähnlichkeitsmaß	Spezies-Grenzwert	Quelle
ANIB <sub>11</sub>	95,25 %	ANIB <sub>11</sub> Logit-Modell (Abbildung 6)
ANIB <sub>12</sub>	94,44 %	ANIB <sub>12</sub> Logit-Modell (Abbildung 6)
ANIm	95,26 %	ANIm Logit-Modell (Abbildung 6)
gANI	AF $\geq$ 0,60 und gANI $\geq$ 96,50 %	Varghese <i>et al.</i> [52]
MLSA	97,00 %	Mulet <i>et al.</i> [25]

## 14. Suche bekannter *B. cereus s.l.* Virulenzfaktoren

### 14.1. Vorhersage homologer Sequenzfamilien

Das Software-Paket Single Linkage Clustering of Sequences (SiLiX, v1.2.9) [396] wurde mit den Standardparametereinstellungen (zzgl. des Parameters '-n' zur Erstellung der .net-Datei) angewendet, um die Menge aller vorhergesagten Proteinsequenzen (Methoden-Abschnitt 9. ) des *B. cereus s.l.* Datensatzes in homologe Sequenzfamilien einzuteilen. Die dafür benötigte jeder-gegen-jeden BLAST-Treffer-Datei wurde im erforderlichen Datei-Format (Parameter '-outfmt 6') mittels Protein-Protein BLAST (blastp v2.2.25+ mit Parametern '-num\_descriptions 10000' und '-num\_alignments 10000') gegen eine entsprechende Protein-

Referenzdatenbank (makeblastdb v2.2.25+) erstellt. Abschließend fand mit Hilfe der High Fidelity Clustering of Sequences (HiFiX, v1.0.5) Software [397] noch eine Unterteilung jeder SiLiX-Sequenzfamilie bestehend aus mindestens zwei Sequenzen (Parameter '-n 2') in HiFiX-Untergruppen statt.

#### 14.2. Untersuchte Virulenzfaktoren

Eine Übersicht der untersuchten *B. cereus s.l.* Virulenzfaktoren mit zugehörigen NCBI Protein Referenzsequenzen geben Tabelle 33 (Seite 213) und Tabelle 34 (Bt-Toxine, Seite 214). Vorbereitend auf die Screening-Prozedur wurden die untersuchten Virulenzfaktoren in unterschiedliche Klassen eingeteilt, nämlich die Enterotoxin-Komplexe (HBL, Nhe), die Hämolysine (CytK, HlyII, HlyIII, HlyA), die Metalloproteasen (InhA1, InhA2, InhA3), die Bt-Toxine (Cry, Cyt, Vip, Sip), die Proteine des Cereulid-Synthetase-Genclusters (*cesHPTABCD*) sowie das Anthrax-Toxin (EF, PA, LF) zuzüglich der drei Komponenten des ABC-Transporters (ATP-Bindeprotein, Permease, Substrat-Bindeprotein) für die Aufnahme von Bikarbonat. Da beim Enterotoxin-Komplex HBL die Existenz eines homologen Komplexes HBL<sub>a</sub> bereits beschrieben wurde [306], enthält die Klasse der Enterotoxin-Komplexe auch Protein Referenzsequenzen für die Einzelkomponenten des homologen Komplexes (HbIC<sub>a</sub>, HbID<sub>a</sub>, HbIA<sub>a</sub>). Ebenso wurde beim Enterotoxin-Komplex Nhe eine seltene homologe Variante Nhe<sub>a</sub> gefunden [26]. Dementsprechend wurden auch hier zugehörige Protein Referenzsequenzen (NheA<sub>a</sub>, NheB<sub>a</sub>, NheC<sub>a</sub>) in die Klasse mit aufgenommen. Die NCBI Protein Referenzsequenzen der 331 Cry, Cyt und Vip Holotypen aus der Klasse der Bt-Toxine konnten mit Hilfe der Website des Komitees für die Nomenklatur von *B. thuringiensis* δ-Endotoxinen (<http://www.btnomenclature.info/>) [272] ausfindig gemacht werden (Stand Feb. 2016). Der Klasse hinzugefügt wurde außerdem noch die bis dato einzig bekannte Sip Holotyp-Sequenz Sip1Aa [273].

#### 14.3. Screening-Prozedur

Zur Ermittlung der An- bzw. Abwesenheit bestimmter Virulenzfaktor CDS in den Genom Assemblierungen der 281 Stämme des *B. cereus s.l.* Datensatzes wurde für jeden Stamm die Menge der vorhergesagten Proteinsequenzen (Methoden-Abschnitt 9. ) als Query-Proteine herangezogen. Jedes Query-Protein wurde mittels Protein-Protein-BLAST (blastp v2.2.25+) gegen eine BLAST Proteindatenbank (makeblastdb v2.2.25+) bestehend aus den Virulenzfaktoren der aktuell betrachteten Virulenzfaktor-Klasse (Methoden-Abschnitt 14.2. ) verglichen. Sofern sich dabei ein

Hit zu einem Virulenzfaktor ( $e\text{-value} \leq 10^{-5}$ ) ergab, wurde dieser gespeichert. Ergaben sich für ein Query-Protein gleich mehrere unterschiedliche Virulenzfaktor-Hits, so wurde nur der beste Hit (höherer Bit-Score) weiter betrachtet. Nach Vergleich aller Query-Proteine eines Stammes auf diese Weise, wurde sofern die besten Hits unterschiedlicher Query-Proteine auf den gleichen Virulenzfaktor fielen wiederum nur der beste Hit (höherer Bit-Score) weiter herangezogen. Damit wurde erreicht, dass jedem beliebigen Virulenzfaktor der aktuell betrachteten Virulenzfaktor-Klasse maximal ein Protein aus demselben Stamm zugeordnet wurde und zwei unterschiedliche Proteine desselben Stammes niemals auf den gleichen Virulenzfaktor fallen konnten. Nachfolgend wurde für jeden Virulenzfaktor über alle Stämme hinweg die Information über die Zugehörigkeit zur homologen SiliX-Familie und HiFiX-Untergruppe (Methoden-Abschnitt 14.1. ) der ihm zugeordneten Proteine gesammelt. So konnten jedem Virulenzfaktor ein oder mehrere SiliX-Familien und HiFiX-Untergruppen zugeteilt werden. Abschließend bildeten die so erhaltenen Virulenzfaktor-assoziierten Familien und Untergruppen den Ausgangspunkt für die eingehende manuelle Untersuchung der sich darin befindlichen proteinkodierenden Sequenzen sowohl auf Aminosäure- wie auch Nukleinsäureebene. Im Zuge der manuellen Inspektion kam der Clustal Omega (v1.2.4) [369-371] Webservice (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) mit Standardparametereinstellungen zur Erstellung von Familien- und Untergruppen-spezifischen multiplen Sequenzalignments zum Einsatz. Auch die zusätzlich zu den multiplen Sequenzalignments erzeugten phylogenetischen Bäume (Neighbour-Joining-Methode ohne Korrektur der Distanz-Werte) und Sequenzähnlichkeitsmatrizen waren Teil der manuellen Überprüfung. Die Analyse der Neighbour-Joining-Bäume erfolgte mit FigTree (v1.4.2, <http://tree.bio.ed.ac.uk/software/figtree/>). Wichtige Erkenntnisse, die sich auf Grundlage der multiplen Clustal Omega Sequenzalignments und Neighbour-Joining-Bäume für ausgewählte homologe Familien ergaben, wurden nochmals mit der MEGA 7 Software (v7.0.26) [373] auf ihre Validität hin überprüft. Für die multiplen Alignments der in diesem Rahmen verwendeten CDS auf Nukleinsäureebene (Gensequenzen) wurde auf die ClustalW Funktionalität (DNA-Alignment mit Standardparametereinstellungen) [372] zurückgegriffen. Die Berechnung der Phylogenien und Visualisierung der Bäume fand entsprechend dem Vorgehen bei der Multi-Locus-Sequenz-Analyse statt (Methoden-Abschnitt 10.4.2. ).

#### **14.4. Visualisierung von An-/Abwesenheitsprofilen**

Die sich auf Grundlage der Virulenzfaktor-assoziierten homologen Familien und Untergruppen ergebenden An- und Abwesenheitsprofile wurden mit der Online-Visualisierungs-Software Tree Of Life (iTOL, v3.6.1) [381] dargestellt.

#### **15. Einfache lineare Regressionsanalyse**

Sämtliche linearen Regressionsanalysen sowie die Berechnung der Bestimmtheitsmaße ( $R^2$ ) wurden mit Gnumeric (v1.10.17) [398] durchgeführt.

## V. ERGEBNISSE

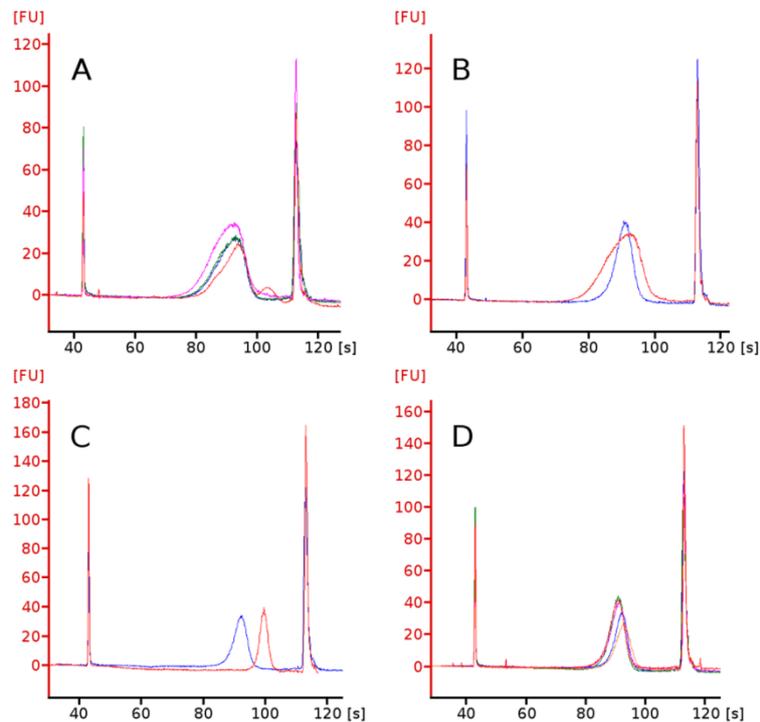
### 1. Optimierung der PCR-freien Probenvorbereitung und Analyse von Faktoren, die die *de novo* Genomassemblierung beeinflussen

#### 1.1. Optimierung des Protokolls zur PCR-freien Herstellung von Illumina Bibliotheken

Die Illumina TruSeq® DNA PCR-free Library Preparation ist eines der am weitesten verbreiteten Protokolle zur Herstellung von paired-end Bibliotheken für die Sequenzierung von Genomen auf den gängigen Illumina Next-Generation-Sequenzierplattformen. Unglücklicherweise können mit diesem Protokoll lediglich Bibliotheken mit kurzen durchschnittlichen Insert-Längen (350 oder 550 bp) präpariert werden. Dies ist unter Berücksichtigung einer maximal sequenzierbaren Read-Länge von 2x300 bp auf der MiSeq Plattform ungünstig, da ein nicht unerheblicher Teil der Read-Paare überlappend sequenziert wird. Eine Erhöhung der Insert-Länge sollte sich jedoch vorteilhaft auf die Assemblierung von paired-end Bibliotheken auswirken, da die Fähigkeit eines Assemblers zur Überbrückung kurzer repetitiver Elemente verstärkt werden sollte. Deshalb wurden im Rahmen dieser Arbeit die Einstellungen für die Fragmentierung der genomischen DNA zu Beginn der Probenvorbereitung angepasst und optimierte Volumenverhältnisse von Resuspension Buffer (RB) zu magnetic Bead Reagent (BR) während den Längenselektionen angewendet (Tabelle 3). Dies ermöglichte die Herstellung von paired-end Bibliotheken, die Unterschiede in ihrer durchschnittlichen Insert-Länge aufwiesen und in die Kategorien IS1 bis IS3 eingeteilt wurden. IS1-Bibliotheken hatten eine durchschnittliche Insert-Länge von ~ 690 bp. IS2- und IS3-Bibliotheken von ~ 990 bp bzw. ~ 1.210 bp (Tabelle 3 und Tabelle 37 auf Seite 224).

Abbildung 1A zeigt die Insert-Längenverteilungen der Bibliotheken, die mit dem TruSeq® DNA PCR-free Library Preparation Protokoll (Kategorie TS) hergestellt wurden. Neben der guten Reproduzierbarkeit weisen alle TS-Bibliotheken eine asymmetrische (linksschiefe) Verteilung auf. Dahingegen können mit dem modifizierten Protokoll Bibliotheken erstellt werden, deren Insert-Längenverteilungen symmetrischer sind und eine geringere Standardabweichung aufweisen, wie es der Vergleich der Bibliotheken Efa\_TS und Efa\_IS1 in Abbildung 1B zeigt. Abbildung 1C verdeutlicht den Längenunterschied von Bibliotheken unterschiedlicher Kategorie.

Gezeigt sind die Bibliotheken Pst\_IS1 und Pst\_IS3, deren durchschnittliche Insert-Längen um ~ 540 bp differieren. Abbildung 1D dient dem Nachweis der hohen Reproduzierbarkeit von Bibliotheken, die mit dem modifizierten Protokoll hergestellt wurden.



**Abbildung 1 Insert-Längenverteilungen nach der zweiten Längenselektion**

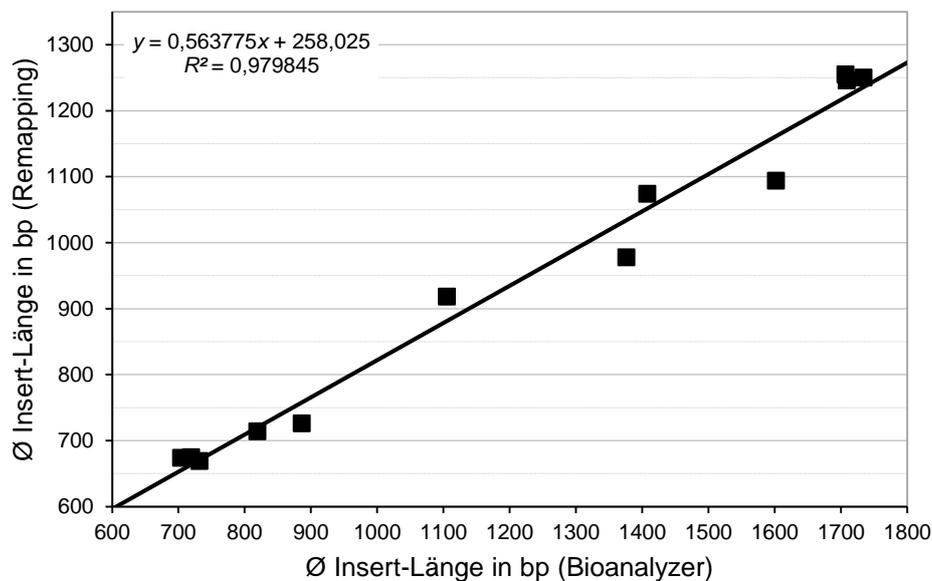
**A** Insert-Längenverteilungen aller Bibliotheken, die mit dem TruSeq® DNA PCR-free Library Preparation Protokoll präpariert wurden. **B** Insert-Längenverteilungen der Bibliotheken Efa\_TS (rot) und Efa\_IS1 (blau). **C** Insert-Längenverteilungen der Bibliotheken Pst\_IS1 (blau) und Pst\_IS3 (rot). **D** Insert-Längenverteilungen aller hergestellten IS1-Bibliotheken.

Diese Abbildung entspricht Abbildung 1 aus der Studie von Huptas *et al.* [342].

Sie unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Die engeren und nahezu symmetrischen Insert-Längenverteilungen ermöglichen eine präzisere Abschätzung der durchschnittlichen Insert-Längen der Bibliotheken. Dennoch stellte sich heraus, dass das Bioanalyzer System dazu tendiert Insert-Längen fehlerhaft wiederzugeben. Ein Remapping der Rohdaten aller sequenzierten IS1-3 Bibliotheken (Tabelle 3) auf deren geschlossene Referenzgenomsequenzen (Tabelle 1) deckte eine sehr starke lineare Korrelation ( $R^2 \sim 0,98$ ,  $p$ -Wert  $\ll 0,001$ ) zwischen den mit dem Bioanalyzer gemessenen und den tatsächlichen durchschnittlichen Insert-Längen auf (Abbildung 2). Dabei überschätzte das Bioanalyzer System die durchschnittliche Insert-Länge von IS1-Bibliotheken im Maximum um 22,0 %. IS3-Bibliotheken wurden schwerwiegend um bis zu 46,4 %

überschätzt (Tabelle 37). Diese Abweichungen scheinen von systematischer Natur zu sein, da zwei unterschiedliche Bioanalyser Maschinen unabhängig voneinander benutzt wurden, um die Insert-Längenverteilungen zu messen. Durch die Berücksichtigung der linearen Regressionsfunktion während der Probenvorbereitung (Methoden-Abschnitt 6.1.7. ), kann die Molarität einer Bibliothek vor ihrer Sequenzierung präziser abgeschätzt werden.



**Abbildung 2 Gemessene und tatsächliche Ø Insert-Längen aller IS1-3 Bibliotheken**

Tatsächliche durchschnittliche Insert-Länge einer Bibliothek in Abhängigkeit von der mit Hilfe des Bioanalyser Systems gemessenen durchschnittlichen Insert-Länge.

Diese Abbildung wurde ins Deutsche übertragen und entspricht Abbildung 2 aus der Studie von Huptas *et al.* [342].

Sie unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Ein weiterer Vorteil des modifizierten Protokolls liegt im reduzierten Verbrauch benötigter Reagenzien. Die Probenvorbereitung von IS1- und IS2-Bibliotheken benötigt nur die Hälfte der Reagenzien, die durch das ursprüngliche Illumina Protokoll verbraucht werden. Für IS3-Bibliotheken muss mehr DNA (~ 150 pMol) auf die Flusszelle geladen werden, um die Bildung adäquater Cluster-Dichten zu gewährleisten. Bei diesen Bibliotheken führt dies zu einem Verbrauch von Reagenzien der dem des originalen Illumina Protokolls entspricht.

## 1.2. Einfluss der Insert-Länge und des GC-Gehaltes einer Bibliothek auf die Qualität der Sequenzierung

Um den Einfluss des GC-Gehaltes und der Insert-Länge einer Bibliothek auf die Qualität der Sequenzierung beurteilen zu können, wurden Stämme mit einem genomischem GC-Gehalt zwischen 35,4 % und 73,0 % ausgewählt (Tabelle 1) und

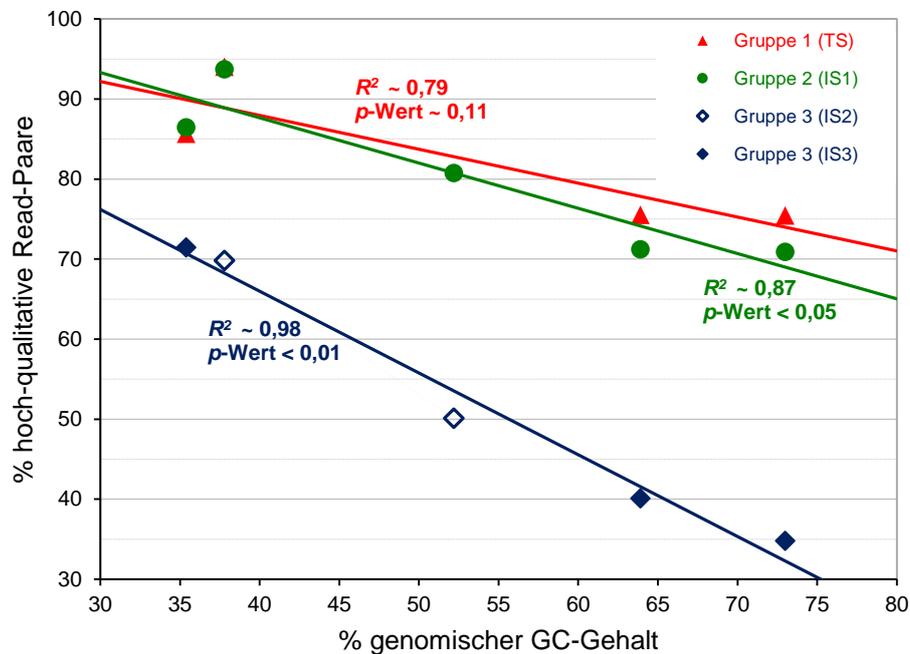
Bibliotheken mit durchschnittlichen Insert-Längen von 610 bp bis 1.250 bp erstellt (Tabelle 3 und Tabelle 38 auf Seite 225). Insgesamt wurden 14 Bibliotheken mit einer Read-Länge von 2x200 bp sequenziert, bevor sie auf eine Read-Länge von 2x189 bp getrimmt wurden. Davon wurden vier Bibliotheken mit dem kommerziellen Illumina Protokoll (Kategorie TS) hergestellt. Fünf der Bibliotheken entsprachen der Kategorie IS1, zwei der Kategorie IS2 und drei der Kategorie IS3. Ein Remapping der Rohdaten gegen ihre entsprechenden Referenzgenomsequenzen ergab, dass TS-Bibliotheken eine ähnliche durchschnittliche Insert-Länge wie ihre IS1 Gegenstücke aufwiesen (Tabelle 38).

In Abbildung 3 ist der GC-Gehalt jeder der 14 Bibliotheken gegen den Prozentsatz ihrer Read-Paare, die die Qualitätskontrolle bestanden hatten, aufgetragen. Die linearen Regressionsanalysen verdeutlichen den generellen Trend, dass die Zunahme des GC-Gehaltes einer Bibliothek eine Abnahme ihrer verbleibenden hoch-qualitativen Read-Paare nach der Qualitätskontrolle zur Folge hat (Gruppen 1-3). Außerdem kann unabhängig vom GC-Gehalt einer Bibliothek festgestellt werden, dass erhöhte durchschnittliche Insert-Längen eine Verminderung des prozentualen Anteils hoch-qualitativer Read-Paare in den Bibliotheken nach sich ziehen (Gruppe 3). In diesem Zusammenhang ist der Unterschied zwischen Gruppe 1 und Gruppe 2 gering, da Bibliotheken der Kategorie TS ähnliche durchschnittliche Insert-Längen wie IS1-Bibliotheken aufweisen.

Der Verlust von Read-Paaren einer Bibliothek aufgrund der Qualitätskontrolle wird begleitet durch eine Verringerung ihrer durchschnittlichen Insert-Länge (Abbildung 4). Die durchschnittliche Insert-Länge der Bibliothek Bce\_IS1 reduzierte sich um ~ 1,2 % von 669 bp der ungefilterten Bibliothek (Abbildung 4A, dunkelgrün) auf 661 bp nach der Qualitätskontrolle (Abbildung 4A, hellgrün). Ähnliches wurde für alle weiteren IS1-Bibliotheken beobachtet (Tabelle 38). Somit wird die Qualität der Sequenzierung bei Bibliotheken der Kategorie IS1 hauptsächlich durch ihren GC-Gehalt bestimmt.

Bei den Bibliotheken Efa\_IS2 und Sen\_IS2 nahm aufgrund der Qualitätskontrolle die durchschnittliche Insert-Länge der gefilterten Bibliothek um ~ 4,5 % (Abbildung 4B) bzw. ~ 6,4 % (Abbildung 4C) ab. Sobald die durchschnittliche Insert-Länge sowie der GC-Gehalt einer Bibliothek weiter ansteigen, prägt sich dieser Effekt noch stärker aus. Zum Beispiel schneidet die GC-reiche Bibliothek Mlu\_IS3 (Abbildung 4D) mit einer Reduktion um ~ 11,8 % wesentlich schlechter ab. Die durch die Qualitätskontrolle induzierte Verlagerung der Insert-Längenverteilung beweist, dass

Read-Paare längerer DNA-Fragmente dazu tendieren von schlechterer Sequenzierqualität zu sein, da diese Paare unverhältnismäßig oft die Filter-Prozedur nicht bestehen. Dieser Effekt ist besonders stark bei Bibliotheken mit hohem GC-Gehalt.



**Abbildung 3 Einfluss der  $\bar{\varnothing}$  Insert-Länge und des GC-Gehaltes auf die Qualität der Sequenzierung**

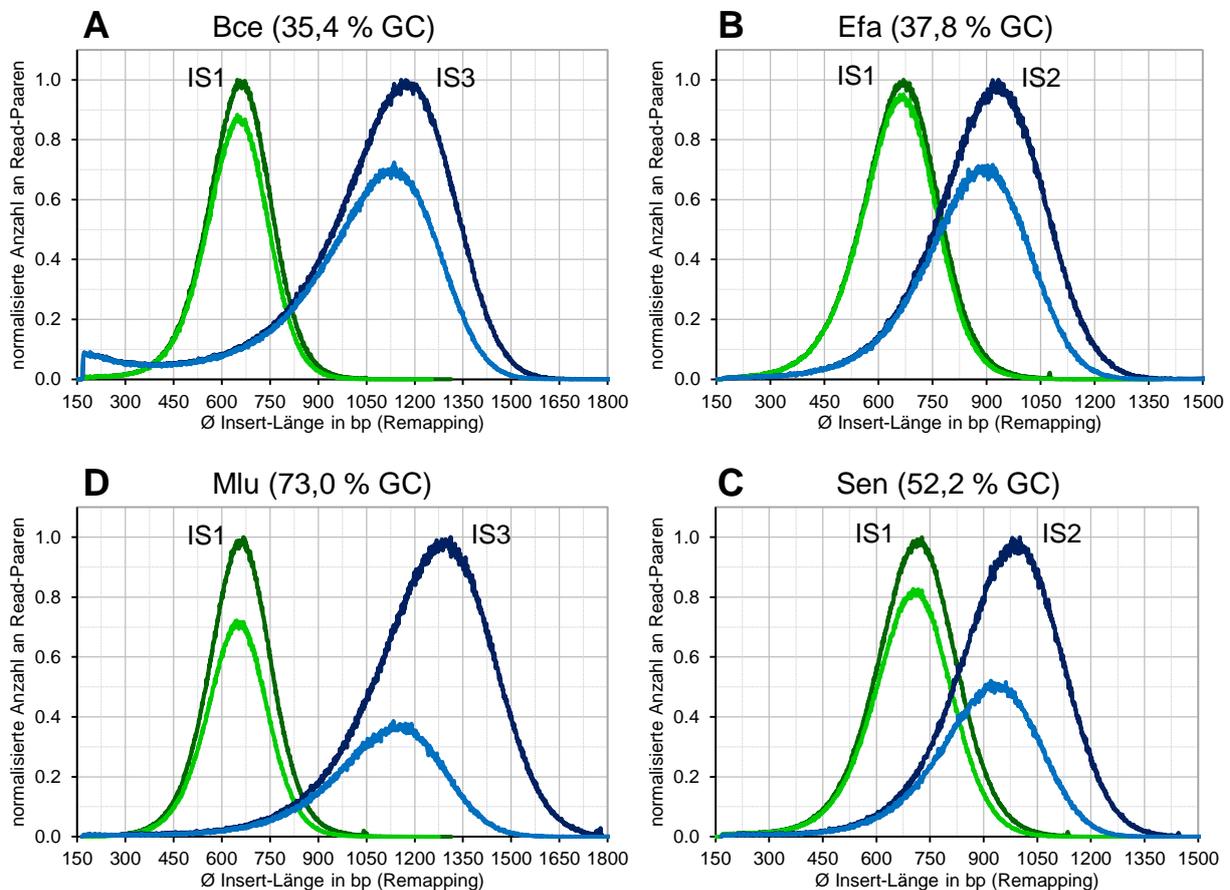
Bibliotheken wurden entsprechend ihrer Kategorie eingruppiert. Gruppe 1 (rot) umfasst alle Bibliotheken, die mit dem ursprünglichen Protokoll (TS) vorbereitet wurden. In Gruppe 2 (grün) sind sämtliche IS1-Bibliotheken zusammengefasst. Gruppe 3 (blau) besteht aus Bibliotheken der Kategorie IS2 oder IS3.

Diese Abbildung wurde ins Deutsche übertragen und entspricht Abbildung 3 aus der Studie von Huptas *et al.* [342].

Sie unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Aus diesem Grund wurde die Auswirkung unterschiedlicher Filter-Einstellungen auf Bibliotheken mit hohem GC-Gehalt im Detail analysiert. Dabei wurden Bibliotheken der Kategorien IS2 bis IS4 für das *Micrococcus luteus* NCTC 2665<sup>T</sup> Genom (73 % GC-Gehalt) hergestellt. Die Sequenzierung dieser Bibliotheken fand mit einer Read-Länge von 2x25 bp (Mlu<sub>50</sub>) statt. Für jede Kategorie wurde die durchschnittliche Insert-Länge der ungefilterten Bibliothek via Remapping ihrer Read-Paare auf die geschlossene Referenzgenomsequenz ermittelt. Anschließend wurde jede ungefilterte Bibliothek einer steigenden Stringenz der Filter-Einstellungen unterworfen und die durchschnittliche Insert-Länge der jeweils verbleibenden hoch-qualitativen Read-Paare via Remapping berechnet (Abbildung 5). Je höher die Kategorie der ungefilterten Bibliothek war, desto ausgiebiger war die Reduktion der durchschnittlichen Insert-Länge nach der Filter-Prozedur. Bei Anwendung der

stringentesten Filter-Einstellung (90;20) betrug der Verlust der Bibliothek mit den kürzesten Insert-Längen (IS2) < 10 % ihrer durchschnittlichen Insert-Länge. Mit ~ 12 % war die Verringerung der IS3-Bibliothek etwas höher, wohingegen sich die durchschnittliche Insert-Länge der IS4-Bibliothek um > 20 % verkleinerte.

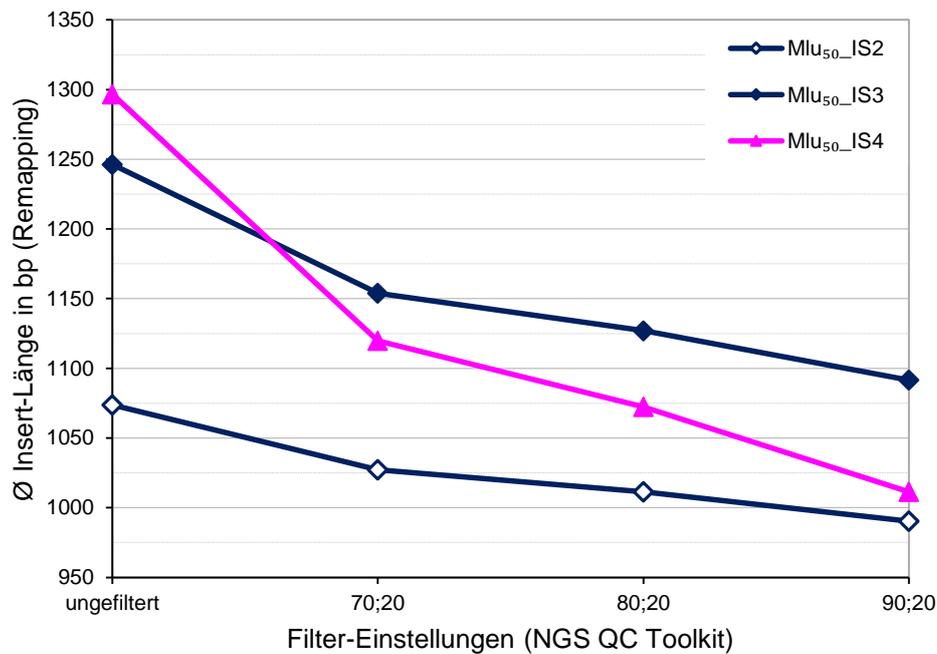


**Abbildung 4 Wechselwirkung zwischen Insert-Länge, GC-Gehalt und der Qualität der Sequenzierung einer Bibliothek**

Darstellung der Insert-Längenverteilungen von IS1-3 Bibliotheken vor und nach der Qualitätskontrolle (80;20). Grüne Verteilungen gehören zu IS1-Bibliotheken. Blaue Verteilungen zu Bibliotheken höherer Kategorie (IS2 oder IS3). Die Verteilungen ungefilterter Bibliotheken sind dunkelfarbig gekennzeichnet. Die gefilterten Bibliotheken hellfarbig. Um die Insert-Längenverteilungen miteinander vergleichbar zu machen, wurde die Anzahl an Read-Paaren durch die maximale Anzahl an Read-Paaren (pro Insert-Länge) der ungefilterten Bibliothek (gleichen Genomes und gleicher Kategorie) normalisiert. **A** Bce\_IS1 und Bce\_IS3. **B** Efa\_IS1 und Efa\_IS2. **C** Sen\_IS1 und Sen\_IS2. **D** Mlu\_IS1 und Mlu\_IS3.

Diese Abbildung wurde ins Deutsche übertragen und entspricht Abbildung 4 aus der Studie von Huptas *et al.* [342].

Sie unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).



**Abbildung 5 Auswirkung der Filter-Einstellungen auf die durchschnittliche Insert-Länge einer Bibliothek**

Durchschnittliche Insert-Längen der Bibliotheken Mlu<sub>50</sub>\_IS2, Mlu<sub>50</sub>\_IS3 und Mlu<sub>50</sub>\_IS4 in Abhängigkeit steigender Filter-Stringenz. Bei der stringentesten Filter-Einstellung müssen beide Reads eines Read-Paares jeweils einen Phred Qualitätswert  $\geq 20$  über mindestens 90 % ihrer Länge aufweisen (90;20), um die Qualitätskontrolle zu bestehen.

Diese Abbildung wurde ins Deutsche übertragen und entspricht Abbildung 5 aus der Studie von Huptas *et al.* [342].

Sie unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

### 1.3. Beeinflussende Faktoren der *de novo* Assemblierung bakterieller Genome

Es gibt eine Vielzahl von Faktoren, die die *de novo* Assemblierung (bakterieller) Genome beeinflussen. Häufig sind entsprechende Faktoren nicht unabhängig voneinander, was es zu einer schwierigen Aufgabe werden lässt den individuellen Einfluss jedes Faktors auf die Genomassemblierung abzuschätzen. Um diese Aufgabe zu lösen wurde eine Strategie herangezogen bei der jeder Faktor parallel zur Ausprägung weiterer Faktoren bewertet wurde (Methoden-Abschnitt 8. ). Die vier in dieser Arbeit betrachteten Faktoren waren: der Assembler, die durchschnittliche Insert-Länge, die Read-Länge sowie die Tiefe der Sequenzierung einer Bibliothek. Die Analysen wurden ausgehend von den Bibliotheken aus Tabelle 3 durchgeführt. Sämtliche Bibliotheken des Stammes *Micrococcus luteus* NCTC 2665<sup>T</sup> wurden allerdings verworfen, da durch das Remapping entsprechender Read-Daten

ersichtlich wurde, dass das sequenzierte Genom zwar sehr ähnlich jedoch nicht identisch zu der geschlossenen NCBI Referenzgenomsequenz aus Tabelle 1 war.

Insgesamt wurden 100 Teil-Bibliothek Triplikate aus dem gesamten Datensatz erzeugt. Daraufhin wurde jedes Mitglied eines Triplikates mit den Assemblern ABySS, Edena, SPAdes und Velvet assembliert, woraus 100 Assemblierungs-Sets pro Assembler entstanden. Die Assemblierungs-Sets wurden mit QUASt gegen ihre Referenzgenomsequenzen (Tabelle 1) validiert. Die für jedes Assemblierungs-Set erhaltenen drei korrigierten NG50 und NGA50 Werte wurden gemittelt. Auf diese Art konnten 400 Validierungs-Einträge (100 pro Assembler) erstellt werden, wobei jeder Eintrag aus einem durchschnittlichen korrigierten NG50 Wert sowie einem durchschnittlichen NGA50 Wert zusammengesetzt war.

Um den besten Assembler zu bestimmen wurden die Validierungs-Einträge zunächst nach ihrem zugrundeliegenden Genom (Stamm) und anschließend nach dem zugrundeliegenden Assembler gruppiert. Um den Einfluss der durchschnittlichen Insert-Länge (Kategorie einer Bibliothek), der Read-Länge und der Tiefe der Sequenzierung untersuchen zu können, fand neben der Eingruppierung nach Genom und Assembler auch noch eine Aufteilung der Validierungs-Einträge in die Ausprägungen des untersuchten Faktors statt. Abschließend wurde für jede Gruppe/Ausprägung der relative Assemblierungs-Wert berechnet. Die Gruppe(n) mit dem höchstem korrigiertem NG50 bzw. NGA50 Wert erhielt(en) den maximalen Assemblierungs-Wert gleich 1. Alle anderen Gruppen wurden relativ zur besten Gruppe bewertet.

SPAdes lieferte eindeutig die besten Assemblierungen (Tabelle 8). Es gibt nur einen Fall (Sen, NGA50) bei dem der maximale relative Assemblierungs-Wert nicht erreicht wurde. Deshalb wird im Folgenden der Einfluss der durchschnittlichen Insert-Länge (Kategorie einer Bibliothek), der Read-Länge sowie der Sequenzierertiefe auf die Leistungsfähigkeit der Assemblierung ausschließlich auf Grundlage der durchgeführten SPAdes Assemblierungen beschrieben. Die Ergebnisse für die Assembler ABySS, Velvet und Edena können Tabelle 39 (Seite 226) bis Tabelle 47 (Seite 230) entnommen werden.

Tabelle 9 zeigt die relativen Assemblierungs-Werte für jedes Genom und jede durchschnittliche Insert-Länge. Zwei Tendenzen werden dabei deutlich. Einerseits weisen Bibliotheken der Kategorie IS1 durchgehend höhere relative Assemblierungs-

Werte als Bibliotheken der Kategorie TS auf. Allerdings gibt es große Unterschiede in der Differenz der Werte bei den verschiedenen Genomen. Im Gegensatz zum GC-reichen Pst Genom, ist der Einfluss der durchschnittlichen Insert-Länge auf die Leistungsfähigkeit der Assemblierung bei den AT-reichen Genomen Bce und Efa wesentlich ausgeprägter. Der zweite Trend liegt darin, dass maximale relative Assemblierungs-Werte am häufigsten bei den erhöhten durchschnittlichen Insert-Längen der Kategorien IS2 und IS3 auftreten. Beide Tendenzen beweisen die Überlegenheit von SPAdes Assemblierungen basierend auf Bibliotheken, die mit dem modifizierten Protokoll hergestellt wurden (IS1-3), über diejenigen des originalen Illumina Protokolls (TS).

**Tabelle 8 Relative Assemblierungs-Werte der verwendeten Assembler**

Assemblierungs-Metrik	Assembler	Genom			
		Bce	Efa	Sen	Pst
korrigierter NG50	SPAdes	1	1	1	1
	ABYSS	0,69	0,87	0,82	0,70
	Velvet	0,71	0,71	0,77	0,94
	Edena	0,63	0,72	0,63	0,73
	<i>max</i>	448.776	381.370	292.477	212.702
NGA50	SPAdes	1	1	0,98	1
	ABYSS	0,93	0,97	1	0,92
	Velvet	0,58	0,82	0,92	0,96
	Edena	0,40	0,66	0,50	0,66
	<i>max</i>	733.293	416.896	405.154	238.037

Für die Genome Bce, Efa und Sen beruht jeder relative Assemblierungs-Wert auf 30 Assemblierungs-Sets.

Für Sen ist die Anzahl der Assemblierungs-Sets für jeden relativen Assemblierungs-Wert gleich 10.

*max* Absoluter Wert in Nukleotiden des maximalen relativen Assemblierungs-Wertes der Spalte.

Diese Tabelle wurde ins Deutsche übersetzt und entspricht Tabelle 3 aus der Studie von Huptas *et al.* [342].

Die Tabelle unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Tabelle 10 umfasst die relativen Assemblierungs-Werte in Bezugnahme auf die Read-Länge. In den meisten Fällen sind für beide Assemblierungs-Metriken die relativen Assemblierungs-Werte maximal oder nahe dem Maximum bei Read-Längen von 175 bp oder 189 bp. Dies impliziert, dass die Stringenz (80;20) der Qualitätskontrolle ausreichend gewesen ist, um einer Verminderung der Leistungsfähigkeit der SPAdes Assemblierungen aufgrund der Anhäufung niedriger Phred Qualitätswerte vorzubeugen. Zudem wird ersichtlich, dass kürzere Read-Längen eine tiefgreifend negative Auswirkung auf die Assemblierung mit SPAdes haben können. Sämtliche Genome mit geringem bis moderatem GC-Gehalt (Bce,

Efa und Sen) weisen bei der kürzesten Read-Länge von 100 bp die kleinsten relativen Assemblierungs-Werte auf.

**Tabelle 9 Relative Assemblierungs-Werte untersuchter durchschnittlicher Insert-Längen (SPAdes)**

Assemblierungs-Metrik	Ø Insert-Länge	Genom			
		Bce	Efa	Sen	Pst
korrigierter NG50	TS	0,54	0,84	n.d.	0,96
	IS1	1	0,95	0,95	0,97
	IS2 oder IS3	0,74	1	1	1
	<i>max</i>	590.346	409.939	300.036	217.864
NGA50	TS	0,53	0,85	n.d.	0,91
	IS1	0,85	0,97	0,96	0,93
	IS2 oder IS3	1	1	1	1
	<i>max</i>	922.992	442.857	405.417	250.853

Grundlage der relativen Assemblierungs-Werte sind ausschließlich SPAdes Assemblierungen.

Für die Genome Bce, Efa und Sen beruht jeder relative Assemblierungs-Wert auf 10 Assemblierungs-Sets.

Für Sen ist die Anzahl der Assemblierungs-Sets für jeden relativen Assemblierungs-Wert gleich 5.

*max* Absoluter Wert in Nukleotiden des maximalen relativen Assemblierungs-Wertes der Spalte.

Diese Tabelle wurde ins Deutsche übersetzt und entspricht Tabelle 4 aus der Studie von Huptas *et al.* [342].

Die Tabelle unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Abschließend wurde der Einfluss der Tiefe der Sequenzierung auf die relativen Assemblierungs-Werte geprüft. Unabhängig von der angewendeten Assemblierungs-Metrik wurden maximale Assemblierungs-Werte immer bei einer Tiefe von 90-fach erreicht (Tabelle 11). Jedoch war ein Anstieg der relativen Assemblierungs-Werte im zweistelligen Prozentbereich nur für Bce auszumachen. Bei den anderen beiden Genomen (Efa und Pst) erbrachte die Verdopplung der Sequenziertiefe von 45- auf 90-fach nur kleine Verbesserungen.

Tabelle 10 Relative Assemblierungs-Werte untersuchter Read-Längen (SPAdes)

Assemblierungs- Metrik	Read-Länge [bp]	Genom			
		Bce	Efa	Sen	Pst
korrigierter NG50	100	0,52	0,68	0,54	0,996
	125	0,82	0,90	0,79	0,995
	150	0,90	0,96	0,81	1
	175	1	1	0,95	0,995
	189	0,91	0,996	1	0,99
	<i>max</i>	<i>541.348</i>	<i>419.532</i>	<i>358.280</i>	<i>213.754</i>
NGA50	100	0,55	0,95	0,71	1
	125	0,75	1	0,86	0,86
	150	0,78	0,97	0,85	0,96
	175	0,996	0,99	0,94	0,98
	189	1	0,999	1	0,97
	<i>max</i>	<i>898.406</i>	<i>424.840</i>	<i>455.293</i>	<i>249.244</i>

Grundlage der relativen Assemblierungs-Werte sind ausschließlich SPAdes Assemblierungen.

Für die Genome Bce, Efa und Sen beruht jeder relative Assemblierungs-Wert auf 6 Assemblierungs-Sets.

Für Sen ist die Anzahl der Assemblierungs-Sets für jeden relativen Assemblierungs-Wert gleich 3.

*max* Absoluter Wert in Nukleotiden des maximalen relativen Assemblierungs-Wertes der Spalte.

Diese Tabelle wurde ins Deutsche übersetzt und entspricht Tabelle 5 aus der Studie von Huptas *et al.* [342].

Die Tabelle unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Tabelle 11 Relative Assemblierungs-Werte untersuchter Sequenzertiefen (SPAdes)

Assemblierungs- Metrik	Tiefe der Sequenzierung [x-fach]	Genom		
		Bce	Efa	Pst
korrigierter NG50	45	0,85	0,94	0,99
	90	1	1	1
	<i>max</i>	<i>485.517</i>	<i>393.034</i>	<i>213.539</i>
NGA50	45	0,78	0,96	0,96
	90	1	1	1
	<i>max</i>	<i>821.716</i>	<i>424.647</i>	<i>242.739</i>

Grundlage der relativen Assemblierungs-Werte sind ausschließlich SPAdes Assemblierungen.

Jeder relative Assemblierungs-Wert beruht auf 15 Assemblierungs-Sets,

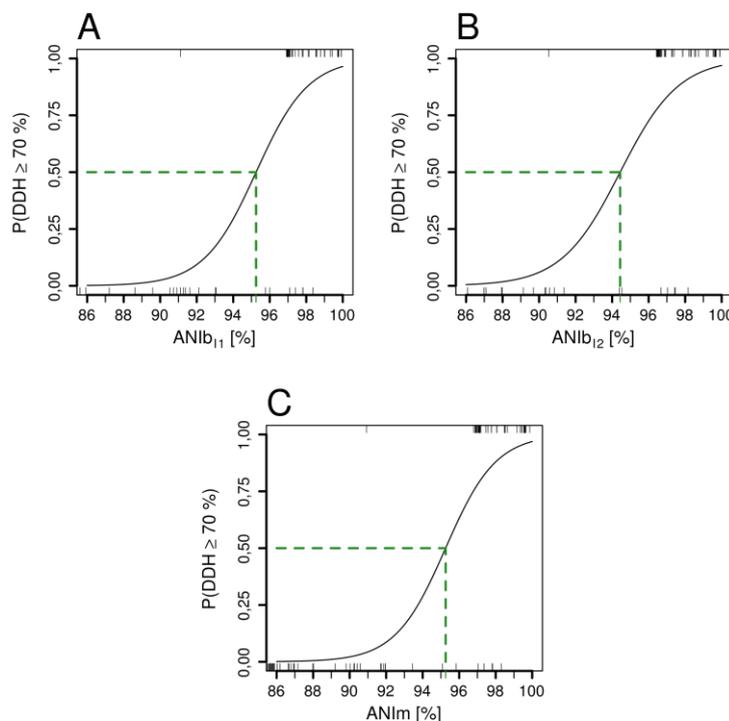
*max* Absoluter Wert in Nukleotiden des maximalen relativen Assemblierungs-Wertes der Spalte.

Diese Tabelle wurde ins Deutsche übersetzt und entspricht Tabelle S5 aus der Studie von Huptas *et al.* [342].

Die Tabelle unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

## 2. Spezies-Grenzwertableitung durch binäre logistische Regression

Auf Basis der DDH Daten von 127 Genompaaren aus 27 prokaryotischen Gattungen wurde für die drei Indices  $ANib_{11}$ ,  $ANib_{12}$  und  $ANIm$  jeweils ein binomiales Logit-Modell geschätzt (Abbildung 6 und Tabelle 12). Erwartungsgemäß handelt es sich bei allen drei Indices um signifikante Einflussgrößen zur Klärung der Frage, ob zwei Genome der gleichen prokaryotischen Art angehören (Tabelle 13). Die durchgehend hohen Pseudo-Bestimmtheitsmaße diverser  $Pseudo-R^2$  Statistiken deuten in jedem der drei Fälle auf eine sehr gute Modellanpassung an die zugrundeliegenden Daten (Tabelle 36) hin, was durch die erreichten AUC-Werte, die allesamt nahe 1 liegen, nochmals bekräftigt wird (Tabelle 14). Wie bei Logit-Modellen üblich wurde der Wendpunkt ( $P(Y = 1) = 0,5$ ) der logistischen Funktion als Schwellenwert für die Klassifikation (Vorhersage) festgesetzt. Je nach Index entsprechen die Spezies-Grenzwerte demnach genomischen Ähnlichkeiten von 95,25 % ( $ANib_{11}$ ), 94,44 % ( $ANib_{12}$ ) bzw. 95,26 % ( $ANIm$ ). Dies ist die erste Arbeit die strikte Spezies-Grenzwerte für eine Reihe von  $ANib/ANIm$  Implementierungen bestimmt hat.



**Abbildung 6** Logit-Modelle der genomischen Ähnlichkeitsmaße  $ANib$  und  $ANIm$

Bei sämtlichen Logit-Modellen wurde der Spezies-Grenzwert auf den Ähnlichkeitswert gesetzt, für den  $P(DDH \geq 70\%) = 0,5$  ist (grün gestrichelte Linie).

**A** Logit-Modell mit Index  $ANib_{11}$  als genomischem Ähnlichkeitsmaß. **B** Logit-Modell mit Index  $ANib_{12}$  als genomischem Ähnlichkeitsmaß. **C** Logit-Modell mit Index  $ANIm$  als genomischem Ähnlichkeitsmaß.

Tabelle 12 Regressionskoeffizienten der Logit-Modelle

Logit-Modell mit Ähnlichkeitsmaß $X$	Regressionskoeffizienten des Logit-Modells		Standardfehler der Regressionskoeffizienten		Effektkoeffizient
	$\hat{\beta}_0$ (Konstante)	$\hat{\beta}_1$	$se(\hat{\beta}_0)$	$se(\hat{\beta}_1)$	$e^{\hat{\beta}_1}$
ANlb <sub>1</sub>	-58,705	0,622	15,867	0,165	1,862
ANlb <sub>2</sub>	-66,523	0,698	17,862	0,185	2,011
ANIm	-69,010	0,725	16,781	0,174	2,064

Tabelle 13 Signifikanz der genomischen Ähnlichkeitsmaße der Logit-Modelle

Logit-Modell mit Ähnlichkeitsmaß $X$	Wald-Test			Likelihood Quotienten-Test				
	$\hat{W}$	Freiheitsgrade	$p$ -Wert (rechtsseitig)	$\hat{D}_0$ (Devianz Nullmodell)	$\hat{D}_1$ (Devianz mit $X$ )	$\hat{D}_0 - \hat{D}_1$ (Differenz der Devianzen)	Freiheitsgrade	$p$ -Wert (rechtsseitig)
ANlb <sub>1</sub>	14,21	1	1,64E-04	156,65	39,81	116,84	1	3,11E-27
ANlb <sub>2</sub>	14,30	1	1,56E-04	156,65	42,09	114,57	1	9,80E-27
ANIm	17,39	1	3,04E-05	156,65	41,91	114,74	1	8,97E-27

Tabelle 14 Gesamtmodell- und Klassifikationsgüte der Logit-Modelle

Logit-Modell mit Ähnlichkeitsmaß $X$	Güte des Gesamtmodells Pseudo- $R^2$ Statistiken			Klassifikationsgüte Receiver-Operating-Characteristic
	Cox & Snells $R^2$	Nagelkerke $R^2$	McFadden $R^2$	AUC
ANlb <sub>1</sub>	0,601	0,849	0,746	0,970
ANlb <sub>2</sub>	0,594	0,839	0,731	0,970
ANIm	0,595	0,839	0,732	0,971

### **3. Taxonomie und Charakterisierung bakterieller Spezies der Gattung *Pseudomonas* sowie der *B. cereus* s.l. Gruppe**

#### **3.1. Speziesabgrenzung innerhalb der Gattung *Pseudomonas***

##### **3.1.1. Der *Pseudomonas* Datensatz**

Der in dieser Arbeit für die Speziesabgrenzung herangezogene *Pseudomonas* Datensatz besteht aus insgesamt 460 Stämmen. Bei 341 Stämmen wurden Genomassemblierungen vom NCBI bezogen (P101-P441 aus Tabelle 28). Für 19 weitere Stämme lagen zwar keine Genomdaten beim NCBI vor, dennoch war es möglich auf dort hinterlegte MLSA Einzelgensequenzen zurückzugreifen (P442-P460 aus Tabelle 30). Dahingegen wurden die Genome von 100 Stämmen eigens *de novo* sequenziert und assembliert (P001-P100 aus Tabelle 2). Dabei wurde auf den im ersten Ergebnis-Abschnitt gemachten Erkenntnissen aufgebaut. So wurden z.B. aufgrund des hohen GC-Gehaltes der *Pseudomonaden* ausschließlich IS1-Sequenzierbibliotheken vorbereitet, die nach erfolgreicher Qualitätskontrolle mit möglichst hoher Read-Länge mit SPAdes assembliert wurden. Tabelle 15 umfasst Details zu den Sequenzierbibliotheken sowie den erreichten Assemblierungs-Statistiken der 100 selbst-sequenzierten *Pseudomonas* Genome.

Tabelle 15 Genomsequenzierungen und Assemblierungen von Stämmen der Gattung *Pseudomonas*

ID	Stammbezeichner	Sequenzierte Read-Länge [bp]	Read-Länge nach Qualitätskontrolle [bp]	Sequenziertiefe [x-fach]	Längstes Contig [bp]	N50	L50	Anzahl Contigs	Assembly Size	WGS Kennung
P001	<i>P. aeruginosa</i> DSM 50071 (T)	2x251	2x240	49	1.322.044	721.732	4	30	6.291.043	JYLC01
P002	<i>P. aeruginosa</i> WS 5022	2x251 & 2x251	2x240 & 2x190	142 (122 & 20)	582.984	345.583	8	68	6.875.522	
P003	<i>P. alcaligenes</i> WS 5013	2x226	2x170	78	1.203.927	710.455	3	10	4.453.840	
P004	<i>P. brenneri</i> DSM 15294 (T)	2x200	2x189	133	742.328	504.582	5	31	5.987.712	
P005	<i>P. brenneri</i> WS 5091	2x251	2x240	100	957.340	490.121	6	36	6.334.542	
P006	<i>P. chengduensis</i> DSM 26382 (T)	2x251	2x240	116	1.012.854	267.193	6	59	5.445.831	
P007	<i>P. chengduensis</i> WS 5019	2x251	2x240	47	855.116	372.704	5	35	5.198.270	
P008	<i>P. chlororaphis</i> (ssp. <i>chlororaphis</i> ) DSM 50083 (T)	2x251 & 2x251	2x240 & 2x150	68 (56 & 12)	1.425.877	841.352	4	25	6.785.600	
P009	<i>P. chlororaphis</i> WS 5014	2x251	2x240	39	2.556.836	1.036.757	2	31	6.787.621	
P010	<i>P. deceptionensis</i> DSM 26521 (T)	2x251	2x240	94	1.862.075	755.170	2	15	5.052.933	JYKX01
P011	<i>P. extremaustralis</i> DSM 17835 (T)	2x226	2x170	41	476.265	187.815	11	106	6.601.321	
P012	<i>P. fluorescens</i> DSM 50090 (T)	2x251	2x240	100	1.089.427	663.730	4	31	6.374.365	
P013	<i>P. fluorescens</i> WS 5037	2x251	2x150	47	1.414.258	392.749	5	48	6.247.579	
P014	<i>P. fragi</i> WS 5025	2x251 & 2x251	2x240 & 2x190	210 (200 & 10)	1.788.461	559.105	3	12	4.993.846	
P015	<i>P. fragi</i> WS 5032	2x251 & 2x251	2x190 & 2x190	76 (19 & 57)	607.252	401.936	5	26	4.930.841	
P016	<i>P. fragi</i> WS 5045	2x251 & 2x251	2x190 & 2x190	81 (18 & 63)	880.959	573.481	4	17	5.025.573	
P017	<i>P. fragi</i> WS 5065	2x251 & 2x251	2x240 & 2x190	223 (208 & 15)	684.259	585.265	4	18	4.907.068	
P018	<i>P. fragi</i> WS 5087	2x251 & 2x251	2x190 & 2x190	76 (16 & 60)	412.135	153.202	12	84	5.387.582	
P019	<i>P. fragi</i> WS 5094	2x251 & 2x251	2x190 & 2x190	62 (13 & 49)	348.477	123.552	14	97	5.333.538	
P020	<i>P. fragi</i> WS 5102	2x251 & 2x251	2x190 & 2x190	115 (23 & 92)	648.123	428.283	6	38	5.496.051	
P021	<i>P. fragi</i> WS 5112	2x251 & 2x251	2x190 & 2x190	69 (14 & 55)	687.771	378.318	6	55	5.456.846	
P022	<i>P. fragi</i> WS 5120	2x251	2x240	107	531.927	169.642	9	96	5.520.514	
P023	<i>P. fragi</i> WS 5123	2x251	2x240	101	2.282.029	761.139	2	13	4.814.352	
P024	<i>P. fragi</i> WS 5124	2x251	2x240	191	813.073	559.756	4	18	5.127.606	
P025	<i>P. fragi</i> WS 5125	2x251	2x240	57	1.706.587	823.015	2	27	4.973.746	
P026	<i>P. gessardii</i> DSM 17152 (T)	2x200	2x189	112	554.093	114.356	18	149	6.508.266	
P027	<i>P. gessardii</i> WS 5049	2x251	2x240	100	404.964	169.792	14	89	6.737.829	
P028	<i>P. grimontii</i> DSM 17515 (T)	2x226	2x170	59	870.260	251.569	10	57	7.040.197	
P029	<i>P. helleri</i> WS 4917 (T)	2x251	2x240	81	623.802	336.668	6	37	5.676.287	JYLD01
P030	<i>P. helleri</i> WS 4995	2x251 & 2x251	2x230 & 2x190	33 (14 & 19)	207.403	93.177	21	164	5.863.021	JYLE01

ID	Stammbezeichner	Sequenzierte Read-Länge [bp]	Read-Länge nach Qualitätskontrolle [bp]	Sequenziertiefe [x-fach]	Längstes Contig [bp]	N50	L50	Anzahl Contigs	Assembly Size	WGS Kennung
P031	<i>P. koreensis</i> DSM 16610 (T)	2x251	2x240	100	1.081.985	689.935	4	27	6.086.647	
P032	<i>P. koreensis</i> WS 5015	2x251	2x240	42	1.479.877	892.370	3	17	6.035.452	
P033	<i>P. "lactis"</i> WS 4992 (iT)	2x251	2x240	48	670.499	392.630	8	42	6.733.097	JYLO01
P034	<i>P. libanensis</i> DSM 17149 (T)	2x251	2x240	44	640.011	326.504	7	43	6.214.234	JYLH01
P035	<i>P. lini</i> DSM 16768 (T)	2x251	2x240	45	993.897	731.568	4	26	6.495.504	JYLB01
P036	<i>P. lundensis</i> DSM 6252 (T)	2x251	2x240	70	393.880	148.083	12	92	4.992.502	JYKY01
P037	<i>P. lundensis</i> WS 5024	2x251	2x150	84	332.854	158.609	13	102	5.219.803	
P038	<i>P. lundensis</i> WS 5026	2x251	2x240	99	639.555	220.333	8	80	5.220.964	
P039	<i>P. lundensis</i> WS 5071	2x251 & 2x251	2x240 & 2x240	110 (87 & 23)	270.091	121.509	14	103	4.803.694	
P040	<i>P. lundensis</i> WS 5078	2x251	2x150	96	697.616	264.864	6	41	4.738.897	
P041	<i>P. lundensis</i> WS 5081	2x251 & 2x226	2x240 & 2x170	77 (17 & 60)	498.402	212.616	9	159	5.215.212	
P042	<i>P. lundensis</i> WS 5088	2x251	2x150	94	699.455	250.066	6	59	5.116.623	
P043	<i>P. lundensis</i> WS 5095	2x251	2x240	39	403.197	108.496	14	124	5.042.919	
P044	<i>P. lundensis</i> WS 5104	2x251	2x240	60	579.070	186.060	9	70	5.171.777	
P045	<i>P. lundensis</i> WS 5110	2x251 & 2x251	2x240 & 2x150	40 (19 & 21)	585.739	196.975	10	80	5.190.045	
P046	<i>P. lundensis</i> WS 5117	2x251	2x240	92	552.719	227.678	7	49	4.908.174	
P047	<i>P. lundensis</i> WS 5118	2x251	2x240	123	528.383	159.291	10	86	5.091.333	
P048	<i>P. lundensis</i> WS 5119	2x251	2x240	132	633.000	184.907	9	95	5.409.286	
P049	<i>P. lundensis</i> WS 5121	2x251	2x240	125	844.669	325.341	5	58	5.007.675	
P050	<i>P. mandelii</i> DSM 17967 (T)	2x251 & 2x251	2x240 & 2x240	46 (10 & 36)	597.965	165.433	11	114	6.988.375	
P051	<i>P. mandelii</i> WS 5114	2x226 & 2x251	2x170 & 2x150	68 (46 & 22)	950.770	690.265	5	35	6.818.060	
P052	<i>P. marginalis</i> DSM 13124 (T)	2x226	2x170	40	651.052	213.631	10	81	7.184.449	
P053	<i>P. nitroreducens</i> DSM 14399 (T)	2x251	2x240	138	1.647.958	855.214	3	50	6.171.316	
P054	<i>P. nitroreducens</i> WS 5012	2x251	2x240	100	784.340	449.658	6	43	6.757.787	
P055	<i>P. orientalis</i> DSM 17489 (T)	2x251	2x240	35	1.019.307	465.046	5	39	6.188.637	JYLM01
P056	<i>P. oryzihabitans</i> DSM 6835 (T)	2x251 & 2x251	2x240 & 2x240	103 (38 & 65)	1.285.244	832.289	3	13	5.038.716	
P057	<i>P. oryzihabitans</i> WS 5017	2x251	2x150	60	1.534.052	1.068.312	2	19	4.995.228	
P058	<i>P. panacis</i> DSM 18529 (T)	2x200	2x189	124	931.687	435.348	5	46	6.801.396	
P059	<i>P. panacis</i> WS 4668	2x200	2x189	95	618.904	404.894	7	37	6.730.220	
P060	<i>P. "paralactis"</i> WS 4672 (iT)	2x251	2x230	39	884.981	652.815	4	37	6.015.212	JYLN01
P061	<i>P. peli</i> DSM 17833 (T)	2x251	2x240	100	2.500.221	2.500.221	1	15	4.511.138	

ID	Stammbezeichner	Sequenzierte Read-Länge [bp]	Read-Länge nach Qualitätskontrolle [bp]	Sequenziertiefe [x-fach]	Längstes Contig [bp]	N50	L50	Anzahl Contigs	Assembly Size	WGS Kennung
P062	<i>P. peli</i> WS 5011	2x251	2x240	100	499.652	216.844	8	88	4.604.264	
P063	<i>P. poae</i> DSM 14936 (T)	2x251	2x240	90	456.710	226.253	9	56	5.883.893	JYLI01
P064	<i>P. poae</i> WS 5103	2x251	2x230	30	592.476	322.592	7	41	5.528.914	
P065	<i>P. protegens</i> WS 5082	2x251	2x150	46	1.794.928	726.802	3	20	6.926.774	
P066	<i>P. proteolytica</i> DSM 15321 (T)	2x200	2x189	121	308.391	129.095	18	107	6.350.040	
P067	<i>P. proteolytica</i> WS 5010	2x251	2x240	106	1.030.439	767.371	4	42	6.231.520	
P068	<i>P. proteolytica</i> WS 5021	2x251	2x230	47	1.393.436	638.072	4	22	6.340.118	
P069	<i>P. proteolytica</i> WS 5027	2x251 & 2x251	2x240 & 2x230	135 (106 & 29)	646.174	199.053	10	57	6.235.040	
P070	<i>P. proteolytica</i> WS 5031	2x251	2x230	67	764.768	274.032	7	47	5.912.257	
P071	<i>P. proteolytica</i> WS 5059	2x251	2x230	32	1.091.327	620.406	4	28	6.195.332	
P072	<i>P. proteolytica</i> WS 5060	2x251 & 2x251	2x230 & 2x190	61 (33 & 28)	306.522	119.534	20	126	6.489.653	
P073	<i>P. proteolytica</i> WS 5079	2x251	2x230	37	1.036.622	473.683	5	35	6.328.935	
P074	<i>P. proteolytica</i> WS 5086	2x251 & 2x226	2x190 & 2x170	59 (16 & 43)	646.450	198.962	10	61	6.148.335	
P075	<i>P. proteolytica</i> WS 5111	2x251 & 2x251	2x190 & 2x190	52 (11 & 41)	374.522	201.692	12	75	6.318.683	
P076	<i>P. proteolytica</i> WS 5126	2x251	2x220	53	362.586	145.597	15	85	6.136.959	
P077	<i>P. proteolytica</i> WS 5127	2x251	2x220	37	782.994	260.396	9	52	5.960.850	
P078	<i>P. proteolytica</i> WS 5128	2x251	2x220	45	873.430	287.694	6	44	5.951.709	
P079	<i>P. proteolytica</i> WS 5145	2x251 & 2x251	2x240 & 2x220	49 (32 & 17)	771.510	285.395	9	67	7.370.329	
P080	<i>P. proteolytica</i> WS 5146	2x251	2x220	33	987.540	541.250	4	37	6.482.218	
P081	<i>P. proteolytica</i> WS 5147	2x251 & 2x251	2x240 & 2x220	90 (60 & 30)	385.700	108.148	18	129	6.421.398	
P082	<i>P. proteolytica</i> WS 5148	2x251 & 2x251	2x240 & 2x220	113 (74 & 39)	438.524	145.786	14	113	6.550.737	
P083	<i>P. psychrophila</i> DSM 17535 (T)	2x251	2x240	33	1.287.586	431.228	4	126	5.334.010	JYKZ01
P084	<i>P. psychrotolerans</i> DSM 15758 (T)	2x251	2x240	48	722.909	324.503	7	40	6.012.013	
P085	<i>P. rhodesiae</i> DSM 14020 (T)	2x200	2x189	133	523.635	239.322	8	82	6.333.130	
P086	<i>P. rhodesiae</i> WS 4669	2x200	2x189	117	670.347	219.939	8	65	5.983.759	
P087	<i>P. rhodesiae</i> WS 5107	2x251	2x150	67	585.474	224.025	8	58	5.961.510	
P088	<i>P. sp.</i> WS 4996	2x251 & 2x251	2x230 & 2x190	70 (25 & 45)	584.982	284.093	9	60	6.730.325	JYLP01
P089	<i>P. sp.</i> WS 5051	2x251	2x240	66	574.148	345.925	6	38	5.142.615	
P090	<i>P. stutzeri</i> WS 5018	2x251	2x150	39	381.357	188.181	9	75	4.803.951	
P091	<i>P. synxantha</i> DSM 18928 (T)	2x251	2x240	60	694.325	326.087	7	64	6.792.979	JYLLJ01
P092	<i>P. taetrolens</i> DSM 21104 (T)	2x251	2x240	110	766.794	596.648	4	24	4.921.904	JYLA01

ID	Stammbezeichner	Sequenzierte Read-Länge [bp]	Read-Länge nach Qualitätskontrolle [bp]	Sequenziertiefe [x-fach]	Längstes Contig [bp]	N50	L50	Anzahl Contigs	Assembly Size	WGS Kennung
P093	<i>P. trivialis</i> DSM 14937 (T)	2x251	2x240	99	738.396	288.356	7	46	5.587.907	JYLK01
P094	<i>P. veronii</i> DSM 11331 (T)	2x200	2x189	91	559.297	176.093	11	95	6.997.867	JYLL01
P095	<i>P. veronii</i> DSM 16272	2x200	2x189	113	276.203	101.893	26	208	7.901.729	
P096	<i>P. veronii</i> WS 4670	2x200	2x189	115	559.018	200.719	10	111	7.064.515	
P097	<i>P. veronii</i> WS 4671	2x200	2x189	102	427.325	161.010	14	125	7.122.147	
P098	<i>P. veronii</i> WS 5113	2x251 & 2x251	2x240 & 2x150	42 (25 & 17)	719.389	217.144	10	77	6.847.663	
P099	<i>P. weihenstephanensis</i> WS 4993 (T)	2x251	2x230	91	855.826	571.018	4	32	4.792.107	JYLF01
P100	<i>P. weihenstephanensis</i> WS 4994	2x251	2x230	84	831.734	427.964	4	47	5.004.386	JYLG01

(T), anerkannter Typstamm nach LPSN [67] (Stand Apr. 2017)

(iT), nicht anerkannter Typstamm, da Arname (noch) nicht in LPSN [67] aufgenommen (Stand Apr. 2017)

### 3.1.2. Aktuelle Taxonomie und erreichte Typstamm Abdeckung

Gemäß LPSN umfasst die Gattung *Pseudomonas* 171 valide beschriebene Spezies (Stand April 2017). Bei insgesamt vier Arten gibt es eine weitere Unterteilung in Unterarten. So setzt sich die Spezies *P. brassicacearum* aus den beiden Unterarten *P. brassicacearum* ssp. *brassicacearum* und *P. brassicacearum* ssp. *neoaurantiaca* zusammen. Die Art *P. cedrina* gliedert sich in die Unterarten *P. cedrina* ssp. *cedrina* und *P. cedrina* ssp. *fulgida*. *P. chlororaphis* besteht aus den vier Subspezies *P. chlororaphis* ssp. *chlororaphis*, *P. chlororaphis* ssp. *aurantiaca*, *P. chlororaphis* ssp. *aureofaciens* und *P. chlororaphis* ssp. *piscium*. *P. oleovorans* wird durch die beiden Unterarten *P. oleovorans* ssp. *oleovorans* und *P. oleovorans* ssp. *lubricantis* repräsentiert. Neben der Einteilung in Unterarten beinhaltet die aktuelle Taxonomie auch eine Reihe von Reklassifizierungen. So gilt *P. pseudoalcaligenes* heute als ein späteres heterotypisches Synonym von *P. oleovorans* [399]. Ähnlich verhält es sich mit *P. multiresinovorans* als späterem heterotypischem Synonym von *P. nitroreducens* [400]. Auch bei den beiden Stämmen *P. chloritidismutans* AW1 [401] und *P. perfectomarina* ATCC 14405 [402] handelt es sich um ehemals anerkannte Typstämme, die der Art *P. stutzeri* angehören [403-405]. Die aktuell geltende *Pseudomonas* Taxonomie nach LPSN (Stand April 2017) ist in Tabelle 48 (Seite 231) zusammengefasst. Anerkannte Typstämme bei denen eine Reklassifizierung als zugehörig zu einer anderen Gattung durch die bisherige Literatur zumindest vorgeschlagen wurde, werden in Tabelle 48 nicht berücksichtigt und gelten in dieser Arbeit von vornherein als nicht zugehörig zur Gattung *Pseudomonas*.

Somit werden 93,57 % (160 von 171) der valide beschriebenen *Pseudomonas* Spezies und 90,91 % (10 von 11) der valide beschriebenen Unterarten durch den *Pseudomonas* Datensatz abgedeckt. Zudem sind 75 % (3 von 4) der reklassifizierten ehemaligen Typstämme mit Verbleib in der Gattung *Pseudomonas* im Datensatz enthalten. Bei alleiniger Betrachtung der Stämme mit zugrundeliegenden Genomdaten (P001-P441 in Tabelle 15 und Tabelle 28) reduziert sich die Anzahl abgedeckter *Pseudomonas* Spezies auf 83,04 % (142 von 171). Die Anzahl abgedeckter Unterarten verringert sich entsprechend auf 72,73 % (8 von 11). Neben den in die LPSN aufgenommenen Arten und Unterarten beinhaltet der Datensatz zusätzlich noch die Genomassemblierungen von sieben Typstämmen (P033, P060, P161, P165, P286, P352 und P433) bereits beschriebener *Pseudomonas* Arten, von

denen davon auszugehen ist, dass sie in (naher) Zukunft in die LPSN aufgenommen werden.

### 3.1.3. MLSA Phylogenie, monophyletische Gruppen und Untergruppen

Um die evolutionären Verwandtschaftsbeziehungen unter den 460 Stämmen des *Pseudomonas* Datensatzes aufzudecken, wurde das weitläufig verwendete MLSA Schema von Mulet *et al.* [25] herangezogen. Dieses Schema macht Gebrauch von partiellen Nukleotidsequenzen der Gene 16S rDNA, *gyrB*, *rpoB* und *rpoD*. Unter Einbeziehung der Außengruppe *C. japonicus* Ueda107<sup>T</sup> wurde ein multiples MLSA Sequenzalignment erzeugt. Auf Basis dieses Alignments konnte durch statistische Modellauswahl das General Time Reversible (GTR) Modell unter Einbeziehung von Ratenvariation (+G) und einem Anteil invariabler Positionen (+I) als das am besten geeignete Nukleotid-Substitutionsmodell ermittelt werden. Unter Annahme dieses Substitutionsmodells wurde ein MLSA Stammbaum mit der ML-Methode erstellt. Aufgrund des Umfangs (461 Stämme) des MLSA Stammbaumes wurde dieser in drei Teilbäume aufgeteilt, welche in Abbildung 7, Abbildung 8 und Abbildung 9 dargestellt sind.

Anhand der topologischen Beschaffenheit des MLSA ML-Stammbaumes wurde die Zugehörigkeit der 460 *Pseudomonas* Stämme zu den 19 bis dato definierten monophyletischen Gruppen und Untergruppen [25, 125, 128] bestimmt. Hinsichtlich der Aufteilung neu hinzugekommener Typstämme konnten 25 valide beschriebene Arten/Unterarten in 13 der bereits geläufigen Gruppen/Untergruppen eingeordnet werden. Darüber hinaus konnten zwei neue Monophyla ausgemacht werden, weshalb vorgeschlagen wird diese als zusätzliche phylogenetische Gruppen mit den Bezeichnungen *P. luteola* und *P. rhizosphaerae* aufzunehmen. Die *P. luteola* Gruppe setzt sich aus den Arten *P. luteola*, *P. duriflava* und *P. zeshuii* zusammen, wohingegen die Spezies *P. rhizosphaerae* und *P. coleopterorum* die *P. rhizosphaerae* Gruppe bilden. Für insgesamt fünf neu hinzugekommene valide beschriebene Arten war eine Zuordnung zu einer Gruppe oder Untergruppe nicht möglich. Die sieben beschriebenen, aber (noch) nicht in die LPSN aufgenommenen Spezies im *Pseudomonas* Datensatz verteilten sich auf fünf der bereits vorhandenen Gruppen und Untergruppen. Der reklassifizierte Typstamm *P. chloritidismutans* AW1 ist erwartungsgemäß Teil der *P. stutzeri* Gruppe. Tabelle 16 gibt eine detaillierte Aufstellung über die Zugehörigkeit aller neu hinzugekommenen Typstämme seit der letzten gleichartigen Analyse von Gomila *et al.* im Jahr 2015 [125].

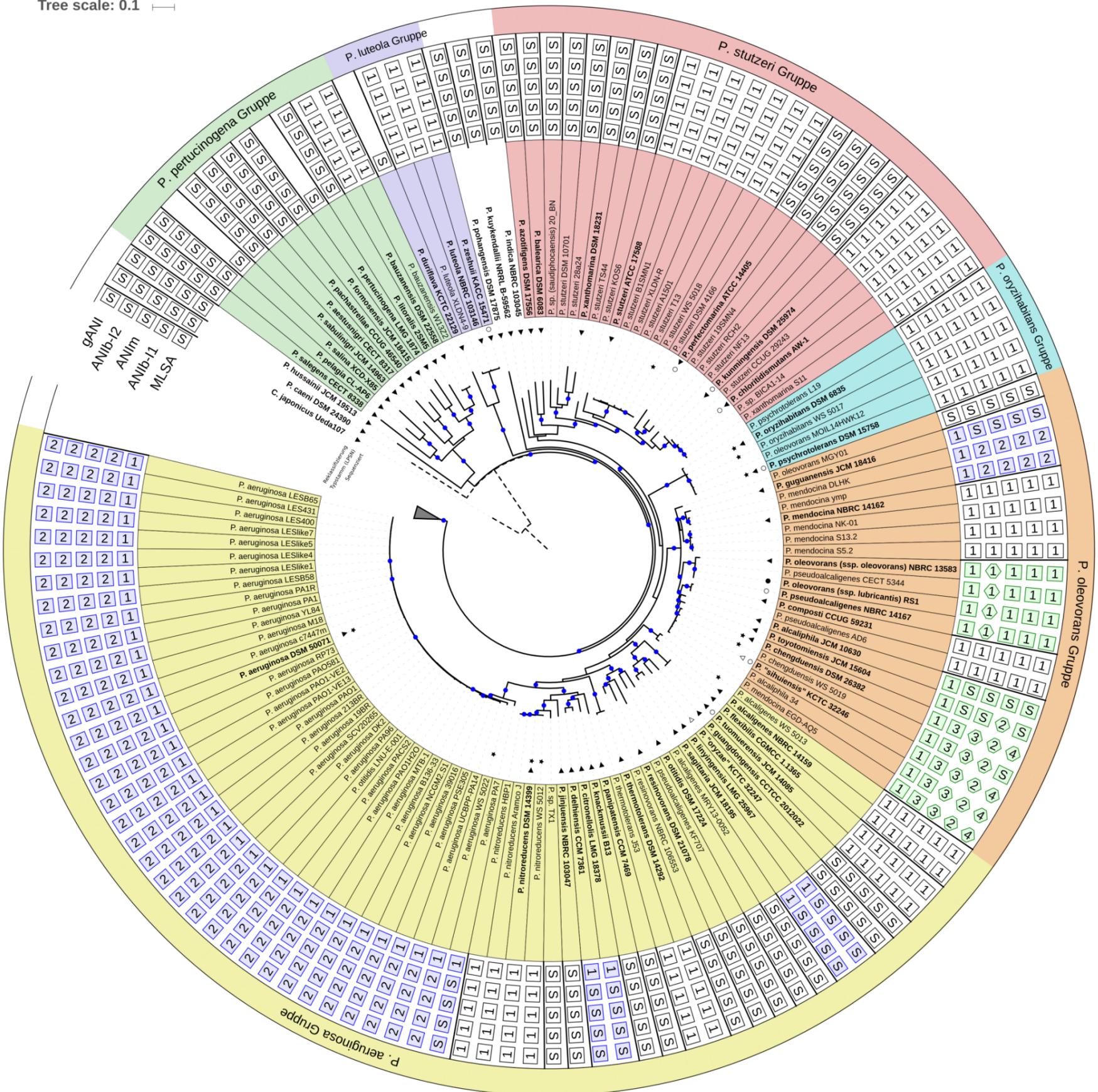
Tabelle 16 *Pseudomonas* Gruppen/Untergruppenzugehörigkeit neuer Typstämme

MLSA ML-Teilbaum	<i>Pseudomonas</i> Gruppe/Untergruppe	Neu hinzugekommener Typstamm (ID)
MLSA ML-Teilbaum 1 (Abbildung 7)	<i>P. pertucinogena</i> Gruppe	<i>P. aestusnigri</i> CECT 8317 (P135)
		<i>P. formosensis</i> JCM 18415 (P231)
		<i>P. salegens</i> CECT 8338 (P347)
	<i>P. luteola</i> Gruppe (neu)	<i>P. salina</i> XCD-X85 (P456)
	<i>P. stutzeri</i> Gruppe	<i>P. zeshuii</i> KACC 15471 (P441)
		<i>P. chloritidismutans</i> AW-1 (P166)
	<i>P. oleovorans</i> Gruppe	<i>P. kunmingensis</i> DSM 25974 (P253)
		<i>P. chengduensis</i> DSM 26382 (P006)
		<i>P. guguanensis</i> JCM 18416 (P245)
		<i>P. oleovorans</i> (ssp. <i>lubricantis</i> ) RS1 (P282)
		<i>P. "sihuiensis"</i> KCTC 32246 (P352)
	<i>P. aeruginosa</i> Gruppe	<i>P. flexibilis</i> CGMCC 1.1365 (P197)
		<i>P. guangdongensis</i> CCTCC 2012022 (P243)
<i>P. linyingensis</i> LMG 25967 (P255)		
<i>P. "oryzae"</i> KCTC 32247 (P286)		
<i>P. sagittaria</i> JCM 18195 (P346)		
ohne Gruppe/Untergruppe	<i>P. hussainii</i> JCM 19513 (P247)	
	<i>P. pohangensis</i> DSM 17875 (P300)	
	<i>P. kuykendallii</i> NRRL B-59562 (P254)	
MLSA ML-Teilbaum 2 (Abbildung 8)	<i>P. straminea</i> Gruppe	<i>P. punonensis</i> CECT 8089 (P311)
	<i>P. rhizosphaerae</i> Gruppe (neu)	<i>P. coleopterorum</i> LMG 28558 (P181)
		<i>P. donghuensis</i> HYS (P191)
		<i>P. guariconensis</i> LMG 27394 (P244)
	<i>P. putida</i> Gruppe	<i>P. soli</i> LMG 27941 (P357)
		<i>P. syringae</i> Gruppe
	<i>P. fragi</i> Untergruppe	<i>P. "cerasi"</i> 58 (P165)
		<i>P. endophytica</i> BSTT44 (P192)
		<i>P. helleri</i> WS 4917 (P029)
		<i>P. "versuta"</i> L10.10 (P433)
		<i>P. weihenstephanensis</i> WS 4993 (P099)
	ohne Gruppe/Untergruppe	<i>P. turukhanskensis</i> IB1.1 (P458)
	MLSA ML-Teilbaum 3 (Abbildung 9)	<i>P. mandelii</i> Untergruppe
<i>P. koreensis</i> Untergruppe		<i>P. granadensis</i> LMG 27940 (P242)
<i>P. chlororaphis</i> Untergruppe		<i>P. chlororaphis</i> (ssp. <i>piscium</i> ) DSM 21509 (P169)
<i>P. gessardii</i> Untergruppe		<i>P. yamanorum</i> LMG 27247 (P440)
		<i>P. "canadensis"</i> 2-92 (P161)
<i>P. fluorescens</i> Untergruppe		<i>P. cedrina</i> (ssp. <i>fulgida</i> ) DSM 14938 (P445)
		<i>P. "lactis"</i> WS 4992 (P033)
		<i>P. "paralactis"</i> WS 4672 (P060)
ohne Gruppe/Untergruppe	<i>P. helmanticensis</i> OHA11 (P448)	

Bei alleiniger Betrachtung der 193 Stämme, welche sowohl in dieser Arbeit als auch schon 2015 Teil der Analyse waren, wird der weitgehende Erhalt der Mitgliederzusammensetzung aller bisherigen Gruppen und Untergruppen deutlich. Lediglich die beiden anerkannten Typstämme *P. cuatrocienegasensis* CIP 109853<sup>T</sup> und *P. composti* CCUG 59231<sup>T</sup> sollten von der *P. aeruginosa* Gruppe in die *P. anguilliseptica* bzw. *P. oleovorans* Gruppe wechseln.

Die überwiegende Mehrheit der monophyletischen Gruppen und Untergruppen (18 von 21) zeichnet sich durch eine hohe Bootstrap-Unterstützung ( $\geq 70\%$ ) aus. Lediglich die *P. stutzeri* Gruppe (52,8 %), die *P. aeruginosa* Gruppe (23,8 %) und die *P. corrugata* Untergruppe (39,2 %) weisen einen geringen bis moderaten Bootstrap-Wert auf.

Tree scale: 0.1



**Abbildung 7 Pseudomonas MLSA ML-Teilbaum 1**

Dargestellt ist die gewurzelte *Pseudomonas* MLSA Phylogenie mit *C. japonicus* Ueda107<sup>T</sup> als Außengruppe (Zentrum). Die phylogenetische Rekonstruktion fand mit der ML-Methode unter Annahme des GTR+G+I Substitutionsmodells statt. Gezeigt ist Teilbaum 1 der Phylogenie. Die Position von Teilbaum 2 (Abbildung 8) ist durch ein graues Dreieck markiert. Zur besseren Darstellung der Astlängen wurden die phylogenetischen Distanzen der Außengruppe sowie des Stammes *P. caeni* DSM 24390<sup>T</sup> künstlich auf 0,3 Substitutionen pro Alignmentposition verkürzt (gestrichelte Äste). Verzweigungen mit hoher Bootstrap-Unterstützung ( $\geq 70\%$  von 500 Replikaten) sind durch einen blauen Kreis gekennzeichnet. Im Rahmen dieser Arbeit *de novo* sequenzierte und assemblierte Genome sind mit einem schwarzen Stern versehen (erster Ring). Darauf folgt die Taxonomie nach LPSN mit Stand vom April 2017 (zweiter Ring). Bei mit einem schwarzen Dreieck versehenen Stämmen handelt es sich um anerkannte Typstämme. Analog stehen schwarze Kreise für die Typstämme valide beschriebener Unterarten. Mit leeren Kreisen markiert sind ehemalige Typstämme, deren Zugehörigkeit zu einer anderen Art bereits durch die LPSN bestätigt wurde. Leere Dreiecke bezeichnen Typstämme beschriebener Arten deren Artnamen noch nicht in die LPSN aufgenommen wurde. Die in dieser Arbeit vorgeschlagenen Reklassifizierungen von Typstämmen finden sich im dritten Ring wieder, wobei die Bedeutung der Symbole des dritten Ringes der des zweiten Ringes entspricht. Der vierte Ring enthält die Bezeichnungen der untersuchten Stämme sowie deren phylogenetische Gruppen/Untergruppenzugehörigkeit. Stämme gleicher Zugehörigkeit sind durch eine gleiche Hintergrundfarbe gekennzeichnet. Weiß hinterlegte Stämme konnten keiner Gruppe bzw. Untergruppe zugeordnet werden. Der Name der jeweiligen Gruppe bzw. Untergruppe ist im letzten Ring aufgeführt. Die vorhergesagten Spezies-Cluster (via MLSA, ANIb<sub>11</sub>, ANIm, ANIb<sub>12</sub> und gANI) sind im fünften Ring kodiert. Eine Erläuterung der Kodierung geben Ergebnis-Abschnitt 3.1.4. und Abbildung 11.

Tree scale: 0.1

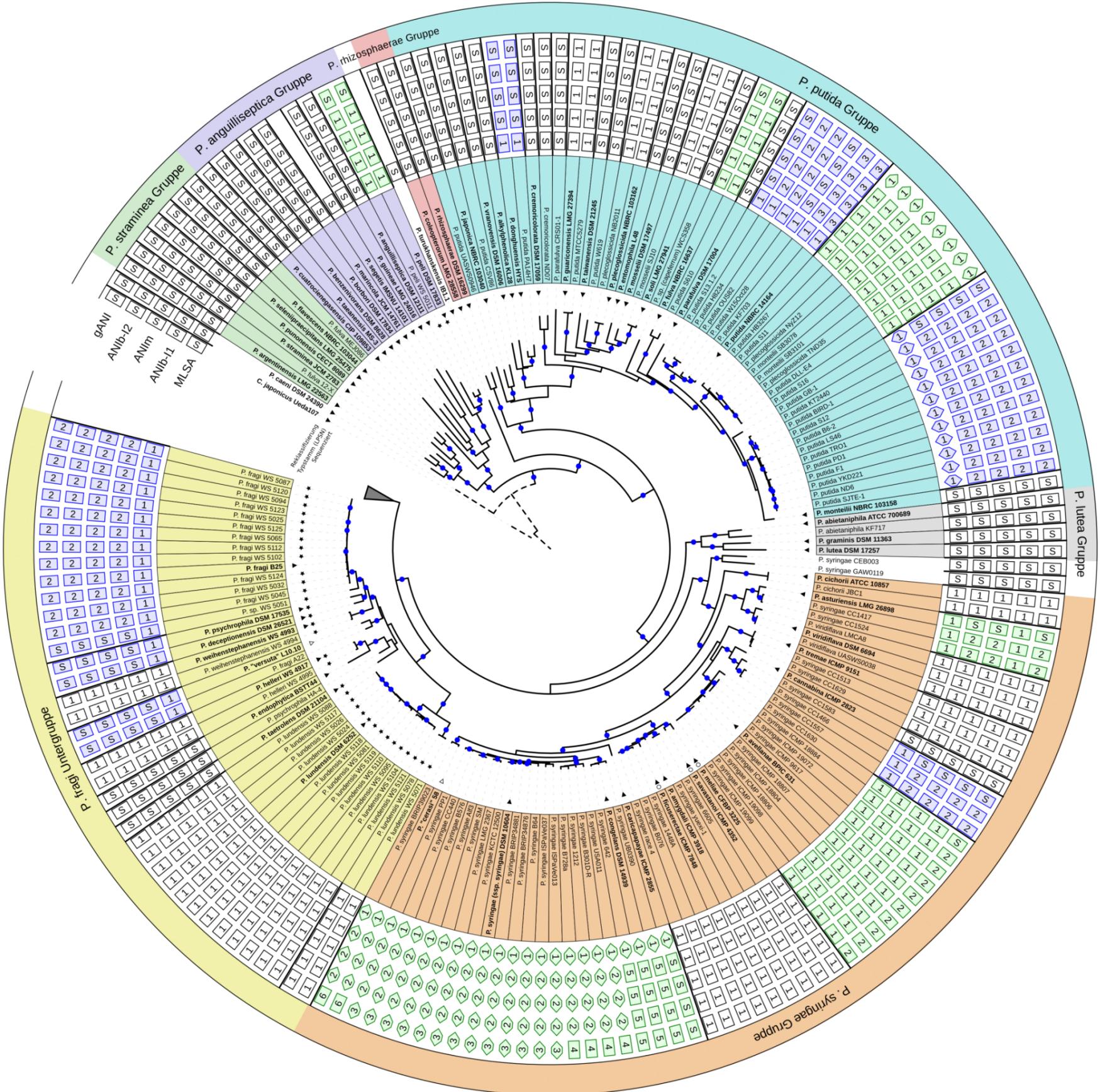


Abbildung 8 *Pseudomonas* MLSA ML-Teilbaum 2

Gezeigt ist Teilbaum 2 der *Pseudomonas* MLSA Phylogenie. Die Position von Teilbaum 3 (Abbildung 9) ist mit einem grauen Dreieck gekennzeichnet. Die phylogenetischen Distanzen der Außengruppe *C. japonicus* Ueda107<sup>T</sup> sowie des Stammes *P. caeni* DSM 24390<sup>T</sup> wurden künstlich auf 0,1 Substitutionen pro Alignmentposition gesetzt (gestrichelte Äste). Ansonsten entspricht der strukturelle Aufbau der Graphik sowie die Bedeutung der darin vorkommenden Symbole denen der in Abbildung 7 gezeigten Graphik von Teilbaum 1.

Tree scale: 0.1

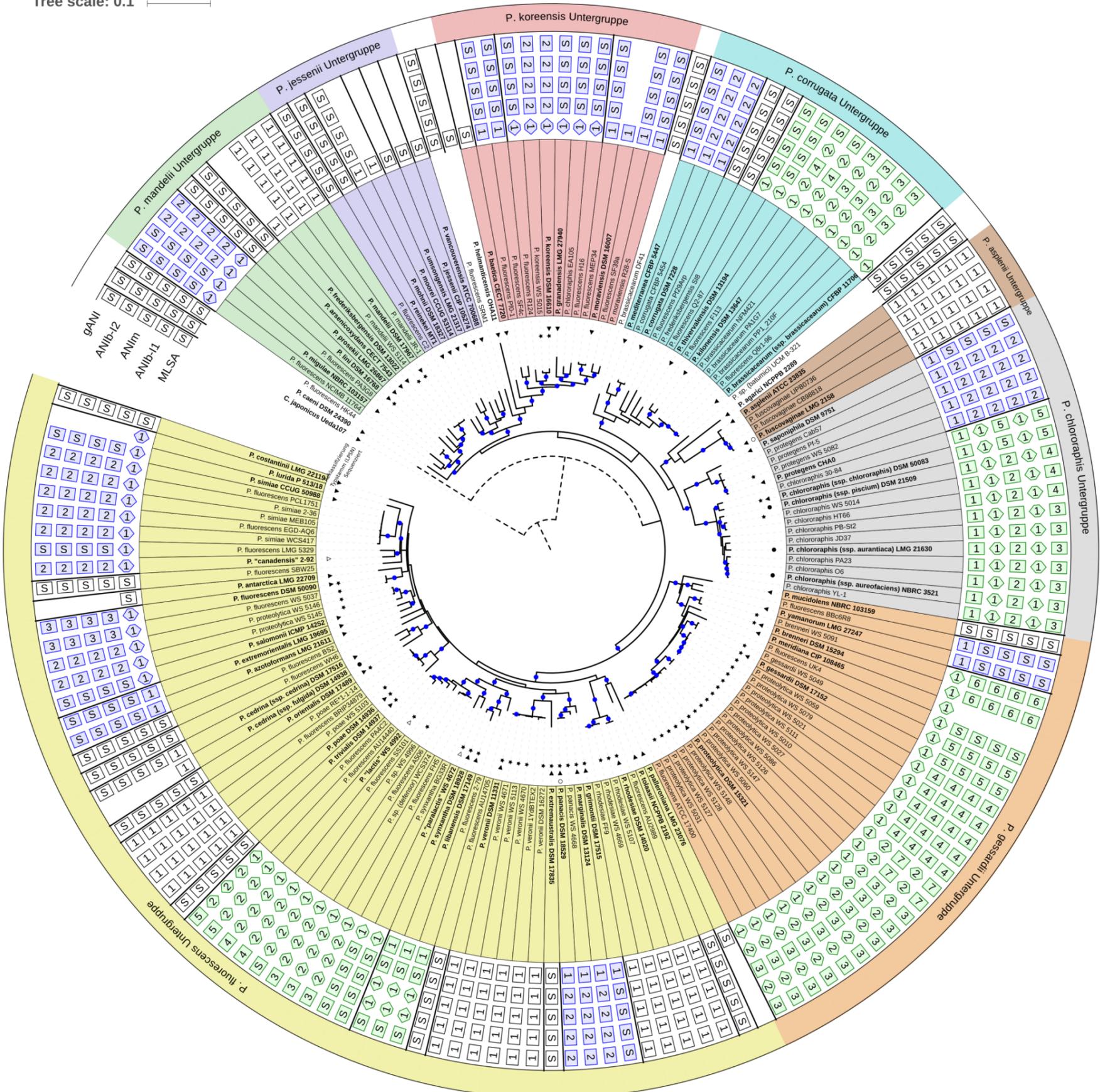


Abbildung 9 *Pseudomonas* MLSA ML-Teilbaum 3

Zu sehen ist Teilbaum 3 der *Pseudomonas* MLSA Phylogenie. Die phylogenetischen Distanzen der Außengruppe *C. japonicus* Ueda107<sup>T</sup> sowie des Stammes *P. caeni* DSM 24390<sup>T</sup> wurden künstlich auf 0,1 Substitutionen pro Alignmentposition gesetzt (gestrichelte Äste). Ansonsten entspricht der strukturelle Aufbau der Graphik sowie die Bedeutung der darin vorkommenden Symbole denen der in Abbildung 7 gezeigten Graphik von Teilbaum 1.

### 3.1.4. Vorhergesagte Spezies-Cluster

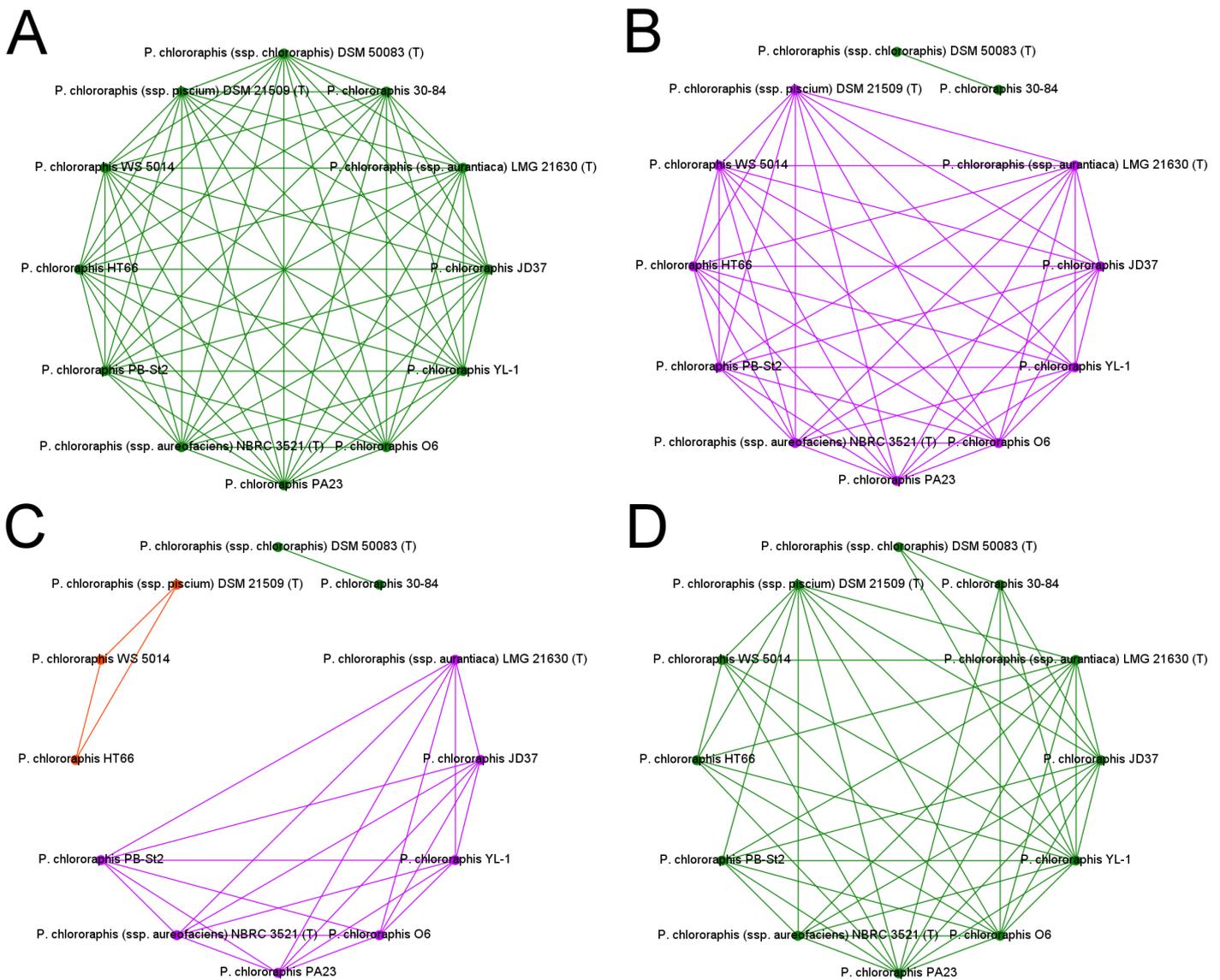
Zur Ermittlung der Spezieszugehörigkeit der 460 Stämme des *Pseudomonas* Datensatzes wurden ein phylogenetischer Index (MLSA Sequenzähnlichkeit) sowie vier phylogenomische Indices (ANlb<sub>11</sub>, ANlb<sub>12</sub>, ANIm und gANI) verwendet. Bei den beiden phylogenomischen Indices ANlb<sub>11</sub> und ANlb<sub>12</sub> handelt es sich lediglich um zwei unterschiedliche Software Implementierungen desselben genomischen Ähnlichkeitsmaßes (ANlb). Zunächst wurden unter Anwendung eines jeder-gegen-jeden Verfahrens alle 105.570 paarweisen MLSA Sequenzähnlichkeiten zwischen den Stämmen berechnet. Im Falle der herangezogenen genomischen Ähnlichkeitsmaße belief sich die Anzahl der errechneten paarweisen Vergleiche auf 97.020 (pro Ähnlichkeitsmaß bzw. ANlb Implementierung), da für 19 *Pseudomonas* Stämme zwar MLSA Sequenzdaten jedoch keine Genomdaten zur Verfügung standen. Anschließend wurden für jeden Index die sich aus dessen paarweisen Vergleichswerten ergebenden Spezies-Cluster bestimmt. Dies wurde durch einen Graphen-basierten Ansatz bewerkstelligt, indem Stämme als Knoten und paarweise Ähnlichkeiten als Kanten zwischen den Knoten dargestellt wurden. Durch Reduktion der Kanten auf diejenigen, deren Vergleichswert größer oder gleich dem Spezies-Grenzwert des aktuell betrachteten Index (Tabelle 7) waren, ergaben sich Index-spezifische Spezies-Cluster in Form von Cliques, Cliques-Gruppen und Singletons (Knoten ohne Kanten).

Als veranschaulichendes Beispiel soll Abbildung 10 dienen. Nach LPSN mit Stand vom April 2017 umfasst die Nomenspezies *P. chlororaphis* vier Unterarten: *P. chlororaphis* ssp. *aurantiaca* (Nakhimovskaya 1948 [406]), *P. chlororaphis* ssp. *aureofaciens* (Kluyver 1956 [407]), *P. chlororaphis* ssp. *chlororaphis* (Guignard & Sauvageau 1894 [408]) und *P. chlororaphis* ssp. *piscium* (Burr *et al.* 2010 [409]). Per definitionem verfügen die Subspezies einer prokaryotischen Nomenspezies über diskriminative phänotypische Eigenschaften [9, 410]. Sobald die genotypische/genomische Kohärenz einer Unterart zu ihrer Nomenspezies nicht mehr aufrechterhalten werden kann (z.B. aufgrund technologischen Fortschritts) repräsentiert die betroffene Unterart eine eigene Nomenspezies [411, 412]. Auf Grundlage der errechneten paarweisen MLSA Sequenzähnlichkeiten gehören zwölf Stämme des *Pseudomonas* Datensatzes der Art *P. chlororaphis* an (Abbildung 10A). Da alle Stämme untereinander MLSA Sequenzähnlichkeiten  $\geq 97$  % aufweisen, ist das zugehörige Spezies-Cluster eine Clique. Im Falle von *P. chlororaphis* stimmt die

Vorhersage des phylogenetischen Index also mit der aktuellen Taxonomie überein, nach der es sich um eine Nomenspezies mit vier Unterarten handelt. Bei Betrachtung des genomischen Ähnlichkeitsmaßes ANIm (Spezies-Grenzwert: 95,26 %) fällt jedoch auf, dass sich die zwölf Stämme in zwei Cliques ohne gemeinsame Knoten aufteilen (Abbildung 10B). Demnach handelt es sich um zwei unterschiedliche Spezies-Cluster und gemäß Regel 50a des International Code of Nomenclature of Bacteria müsste *P. chlororaphis* ssp. *aurantiaca* in den Rang einer Spezies erhoben werden. Unter Anwendung des phylogenomischen Index gANI (Spezies-Grenzwert: 96,5 %), spalten sich die zwölf Stämme sogar in drei Cliques ohne gemeinsame Knoten. Dementsprechend ergeben sich drei unterschiedliche Spezies-Cluster (Abbildung 10C), womit auch die Unterart *P. chlororaphis* ssp. *piscium* in den Rang einer eigenständigen Nomenspezies erhoben werden müsste. Im Gegensatz zu den Cliques der Indices MLSA, ANIm und gANI bilden die beiden ANIb Spezies-Cluster Cliques-Gruppen (Abbildung 10D). In beiden Fällen umfasst die Cliques-Gruppe zwar alle zwölf Stämme, jedoch ist unter den angewendeten Spezies-Grenzwerten (ANIb<sub>11</sub>: 95,25 % und ANIb<sub>12</sub>: 94,44 %) nicht jeder Stamm mit jedem durch eine Kante verbunden. Im Falle von Cliques-Gruppen ergibt sich generell eine Problematik bei der Festlegung und der Differenzierung der Spezieszugehörigkeit einzelner Stämme als auch die Schwierigkeit wie im Zweifelsfall Cliques-Gruppen in unterschiedliche Spezies aufzulösen sind.

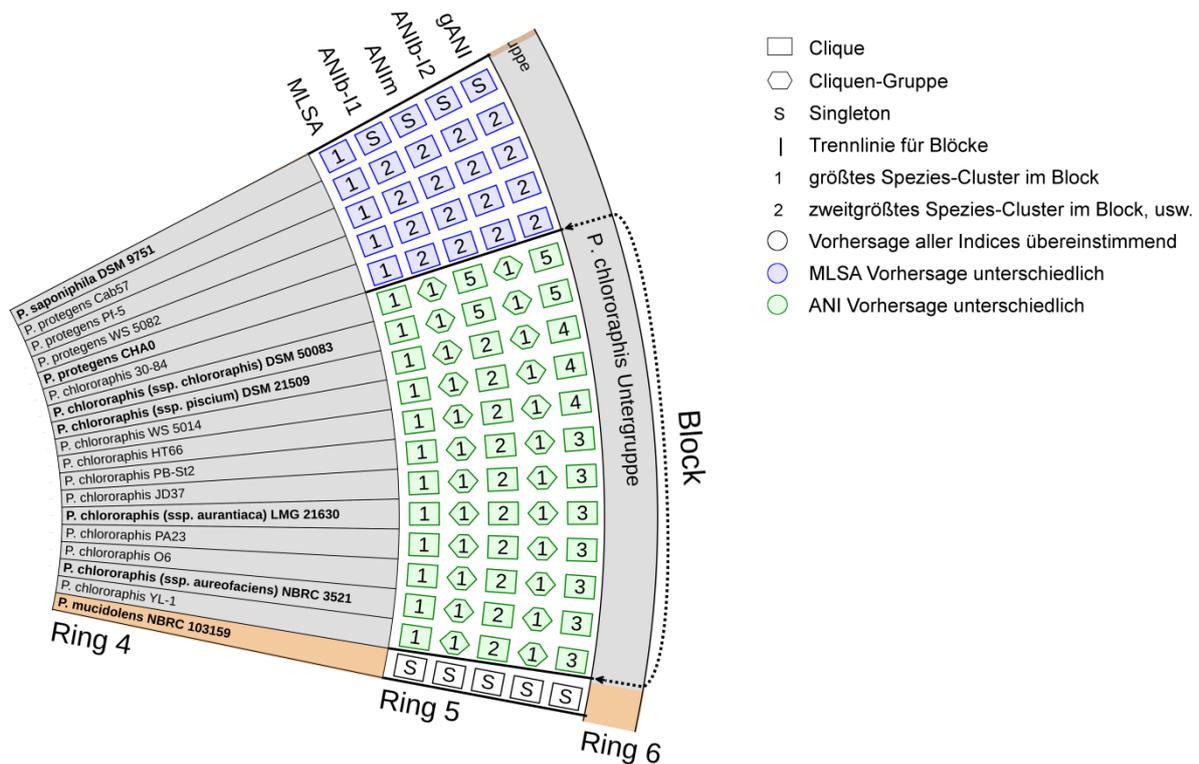
Sämtliche Index-spezifischen Spezies-Cluster, die sich für den *Pseudomonas* Datensatz ergeben, sind im fünften Ring von Abbildung 7, Abbildung 8 und Abbildung 9 kodiert. Die Kodierung wird im Folgenden anhand des Beispiels aus dem vorhergehenden Absatz (Abbildung 10) und in Bezugnahme auf Abbildung 11 erklärt. Mit jeweils zwölf Stämmen beinhalten die Spezies-Cluster der Indices MLSA, ANIb<sub>11</sub> und ANIb<sub>12</sub> die meisten Mitglieder. Damit bilden die zwölf Stämme einen Block, welcher durch zwei Trennstriche innerhalb des fünften Ringes abgegrenzt ist. Die MLSA-Zeile des Blockes ist für jeden der zwölf Stämme mit einem Rechteck und einer darin liegenden "1" gekennzeichnet. Das Rechteck bedeutet, dass die Stämme des MLSA Spezies-Clusters Mitglieder einer Clique sind. Die "1" besagt, dass es sich um ein Block definierendes Cluster (mit der größten Anzahl an Stämmen) handelt. Entsprechend zeigt die ANIb<sub>11</sub>- bzw. ANIb<sub>12</sub>-Zeile Hexagone, um zu verdeutlichen, dass es sich bei den entsprechenden Spezies-Clustern um Cliques-Gruppen handelt. In der ANIm-Zeile sind zehn Stämme durch Rechtecke mit "2"-en versehen, da diese Stämme das zweitgrößte vorhergesagte Spezies-Cluster in Form einer

Clique bilden. Diesem Prinzip folgend stehen die Rechtecke mit einer "3" in der gANI-Zeile für die Vorhersage des drittgrößten Spezies-Clusters in Form einer Clique, usw. Einen Spezialfall bilden Rechtecke mit einem "S". Hierbei handelt es sich um Singletons, die als Cliques mit nur einem Stamm angesehen werden können. Die Rechtecke bzw. Hexagone sind blockweise farblich markiert. Schwarz bedeutet, dass alle Indices in ihrer Vorhersage miteinander übereinstimmen. Blau zeigt an, wenn sich die Vorhersagen der genomischen Ähnlichkeitsmaße untereinander decken die MLSA Vorhersage jedoch davon abweicht. Im Falle von Grün unterscheidet sich die Vorhersage von mindestens einem genomischem Index mit einer beliebigen anderen Vorhersage.



### Abbildung 10 Index-spezifische Spezies-Cluster (*P. chlororaphis*)

**A** Das MLSA Spezies-Cluster entspricht einer Clique und umfasst insgesamt zwölf Stämme (grün). Da die Typstämme (T) der vier Unterarten im Cluster enthalten sind, stimmt die Speziesvorhersage mit der aktuellen Taxonomie nach LPSN überein. **B** Das genomische Ähnlichkeitsmaß ANIm sagt zwei Spezies-Cluster in Form von Cliques vorher. Hier trennt sich der Typstamm *P. chlororaphis* ssp. *chlororaphis* DSM 50083<sup>T</sup> (grün) von den restlichen drei Unterarten (magenta) ab. **C** Bei Anwendung des genomischen Index gANI resultieren sogar drei Spezies-Cluster in Form von Cliques. Dieser Vorhersage zufolge müssten neben *P. chlororaphis* ssp. *chlororaphis* (grün) auch die Unterarten *P. chlororaphis* ssp. *aurantiaca* (magenta) und *P. chlororaphis* ssp. *piscium* (orange) in den Status jeweils eigenständiger Nomenspezies erhoben werden. **D** Bei Betrachtung der ANI<sub>b1</sub> Vergleiche ergibt sich ein Spezies-Cluster in Form einer Cliquen-Gruppe (grün), da innerhalb des Clusters Stammpaare vorhanden sind deren ANI<sub>b1</sub> Werte kleiner dem angewendeten Spezies-Grenzwert sind und damit die entsprechenden Kanten zwischen den Stämmen/Knoten fehlen.



**Abbildung 11 Kodierung Index-spezifischer Spezies-Cluster in Blöcken**

Gezeigt ist ein Ausschnitt des *Pseudomonas* MLSA ML-Teilbaumes 3 aus Abbildung 9. Trennlinien in Ring 5 grenzen unterschiedliche Blöcke voneinander ab. Die Größe der Blöcke wird definiert durch die Anzahl der Stämme im größten vorhergesagten Spezies-Cluster. Rechtecke stehen für Spezies-Cluster in Form von Cliques. Hexagone stehen für Cliquen-Gruppen. Die Nummerierung innerhalb der Rechtecke/Hexagone richtet sich nach der Anzahl der Stämme des jeweiligen Spezies-Clusters. Mit einer "1" versehen sind Stämme, die Mitglieder des größten Clusters sind. Eine "2" tragen Stämme des zweitgrößten Clusters, usw. Ein Sonderfall bilden Rechtecke mit einem "S", was für die Vorhersage eines Singletons steht. Die Vorhersagen der Spezies-Cluster wurden blockweise in drei Kategorien eingeteilt. Schwarz sind Spezies-Cluster deren Vorhersage durch sämtliche Indices deckungsgleich ist. Sofern sich einzig die phylogenetische Vorhersage (MLSA) nicht mit der phylogenetischen Vorhersage (ANIB<sub>11</sub>, ANIB<sub>12</sub>, ANIm und gANI) deckt, sind die entsprechenden Spezies-Cluster blau hinterlegt. Bei grünen Clustern weicht die Vorhersage mindestens eines genomischen Ähnlichkeitsmaßes von einem beliebigen weiteren Index ab.

### 3.1.5. Reklassifizierung von Typstämmen

Für den *Pseudomonas* Datensatz ergeben sich unter Berücksichtigung der aktuellen Taxonomie nach LPSN (Tabelle 48) 12.090 Vergleiche zwischen den vorhandenen Typstämmen mit Genomdaten. Während der Analyse dieser Vergleiche war aufgefallen, dass bei 18 Paaren von Typstämmen die Vorhersageergebnisse sämtlicher Indices (MLSA, ANI<sub>b1</sub>, ANI<sub>b2</sub>, ANI<sub>m</sub>, gANI) untereinander zwar deckungsgleich waren, jedoch der aktuellen Taxonomie entgegenstanden. Aus Gründen der sich daraus ergebenden multiplen Evidenz gegen die aktuelle Taxonomie müsste der Status von insgesamt zwölf Arten geändert werden. Tabelle 17 gibt darüber im Detail Aufschluss. In Übereinstimmung zum Vorschlag einer Reklassifizierung [399] konnte *P. pseudoalcaligenes* als späteres heterotypisches Synonym von *P. oleovorans* ausgemacht werden.

### 3.1.6. Anpassung des MLSA Spezies-Grenzwertes

Um die Güte/Zuverlässigkeit der einzelnen Indices (MLSA, ANI<sub>b1</sub>, ANI<sub>b2</sub>, ANI<sub>m</sub>, gANI) und ihrer Spezies-Grenzwerte (Tabelle 7) bei der Vorhersage der Spezieszugehörigkeit abschätzen zu können, wurde für jeden Index die Anzahl der Typstammkonflikte bezogen auf die Menge aller 12.090 paarweisen Typstammvergleiche mit zugrundeliegenden Genomdaten bestimmt. Grundlage der Analyse war dabei die aktuelle Taxonomie nach LPSN (Tabelle 48) unter Einbeziehung der vorgeschlagenen Änderungen/Reklassifizierungen aus Tabelle 17. Die sich daraus ergebende "neue" Taxonomie wird als Benchmark-Taxonomie bezeichnet. Ein Typstammkonflikt ist definiert als eine der Benchmark-Taxonomie widersprechende Vorhersage der Spezieszugehörigkeit eines Paares von Typstämmen. Bei Betrachtung der Indices als binäre Klassifikatoren entspricht die Anzahl sich ergebender Typstammkonflikte somit der Summe falsch-negativer und falsch-positiver Vorhersagen des jeweiligen Index.

Im Falle des phylogenetischen Index (MLSA) resultieren 24 Typstammkonflikte. Dahingegen sind es bei den phylogenomischen Indices (ANI<sub>b1</sub>, ANI<sub>b2</sub>, ANI<sub>m</sub>, gANI) maximal bis zu fünf Typstammkonflikte (Tabelle 18). In Übereinstimmung mit dieser Beobachtung konnte auch eine überdurchschnittlich hohe Anzahl an blauen Blöcken bei der Bestimmung der *Pseudomonas* Spezies-Cluster (Ergebnis-Abschnitt 3.1.4. ) in den drei MLSA Teilbäumen (Abbildung 7, Abbildung 8 und Abbildung 9) festgestellt werden. Von den insgesamt 35 Blöcken bei denen sich die Vorhersage mindestens eines Index von der eines beliebigen anderen Index unterscheidet (blaue

und grüne Blöcke) sind 22 darauf zurückzuführen, dass die phylogenetische Vorhersage nicht deckungsgleich mit den phylogenomischen Vorhersagen ist (blaue Blöcke). Die hohe Anzahl an Typstammkonflikten und blauen Blöcken macht die Notwendigkeit einer Anpassung des MLSA Spezies-Grenzwertes deutlich. Nach Anhebung des Grenzwertes auf 97,84 % verbleiben vier Typstammkonflikte ohne Einfluss auf die Anzahl richtig-positiver (12.044) und richtig-negativer (22) Vorhersagen unter dem ursprünglichen Grenzwert zu nehmen. Durch den angepassten MLSA Spezies-Grenzwert ist die Güte/Zuverlässigkeit des phylogenetischen Index mit insgesamt vier Typstammkonflikten vergleichbar zur Vorhersagegüte der phylogenomischen Indices.

Tabelle 18 gibt einen Überblick über die (verbleibenden) Typstammkonflikte der einzelnen Indices. Während sich bei den phylogenomischen Indices vor allem Typstammkonflikte zwischen den Unterarten von *P. chlororaphis* ergeben (siehe hierzu auch Abbildung 10), ist es mit Hilfe des phylogenetischen Index trotz Anhebung des Spezies-Grenzwert nicht möglich die Arten *P. congelans* von *P. syringae* und *P. grimontii* von *P. marginalis* als auch von *P. panacis* zu diskriminieren. Zudem kann *P. libanensis* durch die drei Ähnlichkeitsmaße MLSA, ANIm und ANI<sub>b12</sub> nicht von *P. synxantha* unterschieden werden. Schließlich ist auch *P. chengduensis* bei Anwendung des ANI<sub>b12</sub> nicht von *P. toyotomiensis* trennbar.

Tabelle 17 Vorgeschlagene Reklassifizierungen innerhalb der Gattung *Pseudomonas*

<b><i>Pseudomonas</i> Gruppe/Untergruppe</b>	<b>Betroffene Typstämme im Datensatz (ID)</b>	<b>Vorgeschlagene Reklassifizierung</b>
<i>P. luteola</i> Gruppe (MLSA ML-Teilbaum 1, Abbildung 7)	<i>P. luteola</i> NBRC 103146 (P258) <i>P. zeshuii</i> KACC 15471 (P441)	Reklassifizierung von <i>P. zeshuii</i> als späteres heterotypisches Synonym von <i>P. luteola</i> .
<i>P. stutzeri</i> Gruppe (MLSA ML-Teilbaum 1, Abbildung 7)	<i>P. stutzeri</i> ATCC 17588 (P368)	Behandlung von <i>P. perfectomarina</i> und <i>P. chloritidismutans</i> als eigenständige Arten und somit als nicht zugehörig zu <i>P. stutzeri</i> .
	<i>P. perfectomarina</i> ATCC 14405 (P294)	
	<i>P. chloritidismutans</i> AW-1 (P166) <i>P. kunmingensis</i> DSM 25974 (P253)	
<i>P. oryzihabitans</i> Gruppe (MLSA ML-Teilbaum 1, Abbildung 7)	<i>P. oryzihabitans</i> DSM 6835 (P056)	Reklassifizierung von <i>P. psychrotolerans</i> als späteres heterotypisches Synonym von <i>P. oryzihabitans</i> .
	<i>P. psychrotolerans</i> DSM 15758 (P084)	
<i>P. oleovorans</i> Gruppe (MLSA ML-Teilbaum 1, Abbildung 7)	<i>P. chengduensis</i> DSM 26382 (P006)	Keine Anerkennung von <i>P. "sihuiensis"</i> als eigenständige Spezies.
	<i>P. "sihuiensis"</i> KCTC 32246 (P352)	
<i>P. aeruginosa</i> Gruppe (MLSA ML-Teilbaum 1, Abbildung 7)	<i>P. flexibilis</i> CGMCC 1.1365 (P197)	Reklassifizierung von <i>P. tuomuerensis</i> als späteres heterotypisches Synonym von <i>P. flexibilis</i> .
	<i>P. tuomuerensis</i> JCM 14085 (P431)	
<i>P. syringae</i> Gruppe (MLSA ML-Teilbaum 2, Abbildung 8)	<i>P. amygdali</i> ICMP 3918 (P142)	Reklassifizierung von <i>P. meliae</i> , <i>P. savastanoi</i> und <i>P. ficuserectae</i> als spätere heterotypische Synonyme von <i>P. amygdali</i> .
	<i>P. meliae</i> CFBP 3225 (P263)	
	<i>P. savastanoi</i> ICMP 4352 (P350) <i>P. ficuserectae</i> ICMP 7848 (P195)	
<i>P. asplenii</i> Untergruppe (MLSA ML-Teilbaum 3, Abbildung 9)	<i>P. asplenii</i> ATCC 23835 (P146)	Reklassifizierung von <i>P. fuscovaginae</i> als späteres heterotypisches Synonym von <i>P. asplenii</i> .
	<i>P. fuscovaginae</i> LMG 2158 (P239)	
<i>P. fluorescens</i> Untergruppe (MLSA ML-Teilbaum 3, Abbildung 9)	<i>P. marginalis</i> DSM 13124 (P052)	Reklassifizierung von <i>P. panacis</i> als späteres heterotypisches Synonym von <i>P. marginalis</i> .
	<i>P. panacis</i> DSM 18529 (P058)	

Tabelle 18 Typstammkonflikte innerhalb des *Pseudomonas* Datensatzes nach Abgleich mit der Benchmark-Taxonomie

Betroffene Typstämme im Datensatz (ID)	MLSA [%] (Grenzwert: 97,84 %)	ANIB <sub>11</sub> [%] (Grenzwert: 95,25 %)	ANIm [%] (Grenzwert: 95,26 %)	ANIB <sub>12</sub> [%] (Grenzwert: 94,44 %)	gANI [%] (Grenzwert: 96,50 %)
<i>P. chengduensis</i> DSM 26382 (P006)	---	---	---	94,50	---
<i>P. toyotomiensis</i> JCM 15604 (P429)	---	---	---	---	---
<i>P. chlororaphis</i> (ssp. <i>aurantiaca</i> ) LMG 21630 (P167)	---	95,02	94,94	---	95,43
<i>P. chlororaphis</i> (ssp. <i>chlororaphis</i> ) DSM 50083 (P008)	---	---	---	---	---
<i>P. chlororaphis</i> (ssp. <i>aurantiaca</i> ) LMG 21630 (P167)	---	---	---	---	96,00
<i>P. chlororaphis</i> (ssp. <i>piscium</i> ) DSM 21509 (P169)	---	---	---	---	---
<i>P. chlororaphis</i> (ssp. <i>aureofaciens</i> ) NBRC 3521 (P168)	---	94,48	94,94	---	95,50
<i>P. chlororaphis</i> (ssp. <i>chlororaphis</i> ) DSM 50083 (P008)	---	---	---	---	---
<i>P. chlororaphis</i> (ssp. <i>aureofaciens</i> ) NBRC 3521 (P168)	---	---	---	---	96,03
<i>P. chlororaphis</i> (ssp. <i>piscium</i> ) DSM 21509 (P169)	---	---	---	---	---
<i>P. chlororaphis</i> (ssp. <i>chlororaphis</i> ) DSM 50083 (P008)	---	95,10	94,75	94,42	95,35
<i>P. chlororaphis</i> (ssp. <i>piscium</i> ) DSM 21509 (P169)	---	---	---	---	---
<i>P. congelans</i> DSM 14939 (P183)	98,58	---	---	---	---
<i>P. syringae</i> (ssp. <i>syringae</i> ) DSM 10604 (P380)	---	---	---	---	---
<i>P. grimontii</i> DSM 17515 (P028)	98,55	---	---	---	---
<i>P. marginalis</i> DSM 13124 (P052)	---	---	---	---	---
<i>P. grimontii</i> DSM 17515 (P028)	98,17	---	---	---	---
<i>P. panacis</i> DSM 18529 (P058)	---	---	---	---	---
<i>P. libanensis</i> DSM 17149 (P034)	98,72	---	95,67	94,99	---
<i>P. synxantha</i> DSM 18928 (P091)	---	---	---	---	---

### 3.1.7. Anzahl neuer Spezies

Hinsichtlich der *Pseudomonas* Benchmark-Taxonomie (Tabelle 48 unter Einbeziehung der Vorschläge aus Tabelle 17) besteht die Gattung *Pseudomonas* aus 164 Arten sowie sechs weiteren (noch) nicht anerkannten Typstämmen (P033, P060, P161, P165, P286 und P433 in Tabelle 15 bzw. Tabelle 28). Die hiesige Extrapolation der Anzahl neuer *Pseudomonas* Spezies geht jedoch nur von 159 (MLSA<sup>159</sup>) bzw. 141 (MLSA<sup>141</sup>, ANI<sub>b1</sub>, ANI<sub>m</sub>, ANI<sub>b2</sub> und gANI) bekannten *Pseudomonas* Arten aus (Tabelle 19), da für insgesamt elf valide beschriebene Arten nicht ersichtlich ist, ob sie durch den *Pseudomonas* Datensatz abgedeckt werden (keine Genom- bzw. MLSA-Daten der entsprechenden anerkannten Typstämme vorhanden). Weiterhin sind für 18 anerkannte Typstämme zwar MLSA- jedoch keine Genomdaten vorhanden. Tabelle 19 zeigt, dass je nach betrachtetem Index die geschätzte Anzahl neuer *Pseudomonas* Genomospezies zwischen  $\geq 57$  (MLSA) und  $\geq 81$  (gANI) liegt.

Tabelle 19 Anzahl bekannter und potentieller neuer *Pseudomonas* Spezies

Ähnlichkeitsmaß	Gesamtanzahl vorhergesagter			Anzahl bekannter Spezies in Form von			Anzahl potentieller neuer Spezies in Form von		
	Singletons	Cliquen	Cliquen-Gruppen	Singletons	Cliquen	Cliquen-Gruppen	Singletons	Cliquen	Cliquen-Gruppen
MLSA <sup>159</sup> (Spezies-Grenzwert: 97,84 %)	150	56 (260 Stämme)	4 (50 Stämme)	101	51 (228 Stämme)	7 (50 Stämme)	49	8 (32 Stämme)	≥ 0 <sup>a</sup>
MLSA <sup>141</sup> (Spezies-Grenzwert: 97,84 %)	134	55 (257 Stämme)	4 (50 Stämme)	85	49 (225 Stämme)	7 (50 Stämme)	49	8 (32 Stämme)	≥ 0 <sup>a</sup>
ANlb <sub>1</sub>	150	52 (231 Stämme)	6 (60 Stämme)	91	43 (195 Stämme)	7 (60 Stämme)	59	9 (36 Stämme)	≥ 0 <sup>a</sup>
ANlm	147	58 (259 Stämme)	4 (35 Stämme)	89	46 (207 Stämme)	6 (35 Stämme)	58	12 (52 Stämme)	≥ 0 <sup>a</sup>
ANlb <sub>2</sub>	145	52 (231 Stämme)	7 (65 Stämme)	87	44 (197 Stämme)	10 (65 Stämme)	58	8 (34 Stämme)	≥ 0 <sup>a</sup>
gANI	159	60 (257 Stämme)	3 (25 Stämme)	94	45 (204 Stämme)	2 (17 Stämme)	65	15 (53 Stämme)	≥ 1 <sup>a</sup> (8 Stämme)

<sup>159</sup>, Statistik basierend auf 159 bekannten Arten aus der Benchmark-Taxonomie mit zugrundeliegenden MLSA-Daten.

<sup>141</sup>, Statistik basierend auf 141 bekannten Arten aus der Benchmark-Taxonomie mit zugrundeliegenden Gendomdaten.

<sup>a</sup>, Aufgrund der fehlenden Auflösung von Cliquen-Gruppen in einzelne Spezies wurde die Anzahl neuer Arten in Cliquen-Gruppen mit mindestens einem Typstamm standardmäßig auf ≥ 0 festgelegt. Bei Cliquen-Gruppen ohne Typstamm entsprechend auf ≥ 1.

### 3.2. Speziesabgrenzung innerhalb der *B. cereus s.l.* Gruppe

#### 3.2.1. Der *B. cereus s.l.* Datensatz

Die überwiegende Mehrheit der 281 Genomsequenzierungen des *B. cereus s.l.* Datensatzes (B009-B281 aus Tabelle 31) wurde über das NCBI bezogen. Daneben wurden acht Stämme eigens *de novo* sequenziert und assembliert (B001-B008 aus Tabelle 2), um die geringe Gesamtanzahl von Isolaten bestimmter Spezies (z.B. *B. cytotoxicus*) zu erhöhen. Tabelle 20 (Seite 107) gibt Aufschluss über die angefertigten Sequenzierbibliotheken und die erzielten Assemblierungs-Statistiken der selbst-sequenzierten Stämme.

Nach LPSN (Stand November 2017) umfasst die *B. cereus s.l.* Gruppe die neun valide beschriebenen Arten *B. anthracis* [162], *B. cereus s.s.* [161], *B. cytotoxicus* [167], *B. mycoides* [164], *B. pseudomycoides* [165], *B. thuringiensis* [163], *B. toyonensis* [168], *B. weihenstephanensis* [166] und *B. wiedmannii* [169]. Sieben der Arten sind mit der Genomsequenzierung ihres anerkannten Typstammes im Datensatz enthalten. Im Falle von *B. anthracis* ist die Präsenz des Typstammes nicht zwingend erforderlich, da es sich hier um eine hoch klonale Gruppe mit sehr geringer Genomdiversität handelt [240, 352]. Somit ist lediglich die Spezies *B. wiedmannii* nicht durch ihren anerkannten Typstamm im Datensatz vertreten. Durch nachträgliche ANI-Screenings (ANIB<sub>11</sub>, ANIm, gANI) der *B. wiedmannii* FSL W8-0169<sup>T</sup> Genomsequenzen (NCBI WGS Kennung LOBC01) gegen den *B. cereus s.l.* Datensatz war es jedoch möglich die zur Art *B. wiedmannii* gehörigen Stämme im Datensatz ausfindig zu machen. Neben den neun validen Arten umfasst der *B. cereus s.l.* Datensatz auch die Typstämme dreier beschriebener Arten ohne bisherige Aufnahme in die LPSN. Diese sind *B. "bingmayongensis"* [172], *B. "gaemokensis"* [170] und *B. "manliponensis"* [171].

#### 3.2.2. MLSA Phylogenie, vorhergesagte Spezies-Cluster und Anzahl neuer Spezies

Die Rekonstruktion der Phylogenie der 281 Stämme des *B. cereus s.l.* Datensatzes basiert auf den vollständigen proteinkodierenden Sequenzen der sechs Housekeeping-Gene *adk*, *ccpA*, *glpF*, *panC*, *pta* und *pycA*. Ursprünglich war die Verwendung eines bereits etablierten *B. cereus s.l.* MLSA-Schemas [26, 367], das sich zudem auf das Gen *glpT* stützt, angedacht. Allerdings fehlt das Gen in sechs Stämmen (B173, B186, B188, B198, B207 und B218 aus Tabelle 31) und ist in weiteren fünf Stämmen (B026, B051, B147, B254 und B259 aus Tabelle 31) des

Datensatzes nur partiell enthalten, was seine MLSA-Tauglichkeit grundsätzlich in Frage stellt. Abbildung 12 und Abbildung 13 zeigen die beiden Teilbäume der *B. cereus s.l.* Phylogenie, die mit Hilfe der ML-Methode unter Annahme des GTR+G+I Substitutionsmodells und *B. subtilis* ATCC 6051<sup>T</sup> als Außengruppe rekonstruiert wurde. Neben der Phylogenie sind auch die vorhergesagten *B. cereus s.l.* Spezies-Cluster (letzter Ring) in den beiden Abbildungen dargestellt. Die Vorhersage der Spezies-Cluster basiert auf jeweils 97.020 paarweisen Stammvergleichen (jeder-gegen-jeden) mit den drei phylogenomischen Indices ANI<sub>b1</sub>, ANI<sub>m</sub> und gANI. Eine detaillierte Erläuterung über die Kodierung Index-spezifischer Spezies-Cluster wurde bereits in Ergebnis-Abschnitt 3.1.4. sowie Abbildung 11 gegeben.

Mit Hilfe der sich auf phylogenomischer Basis (ANI<sub>b1</sub>, ANI<sub>m</sub>, gANI) ergebenden Spezies-Cluster konnte der *B. cereus* Datensatz in 20 Blöcke unterteilt werden (Abbildung 12 und Abbildung 13). Jeder der 20 Blöcke entspricht einer monophyletischen Gruppe in der MLSA ML-Phylogenie. Zudem weisen mit Einschränkung von Block 14 (65,4 %) alle Blöcke eine durchgehend hohe Bootstrap-Unterstützung ( $\geq 70$  %) auf. Für 16 der 20 Blöcke ist die Vorhersage der Spezies-Cluster mit Hilfe aller drei phylogenomischen Indices deckungsgleich (schwarze Cluster). Da es sich bei den besagten Spezies-Clustern ausschließlich um Cliques (und Singletons) handelt, kann die dem jeweiligen Block entsprechende monophyletische Gruppe von Stämmen als wohl-definierte Spezies betrachtet werden. In diesem Zusammenhang bezeichnet der Ausdruck "wohl-definierte Spezies" eine Teilmenge von Stämmen (z.B. die drei Stämme aus Block 11) aus einer endlichen Menge betrachteter Stämme (hier die 281 Stämme des *B. cereus s.l.* Datensatzes) für die es keine Unsicherheit/Zweifel über die Zugehörigkeit zur selben Spezies sowohl nach innen (alle paarweisen Stammvergleiche innerhalb der Teilmenge sind größer oder gleich den angewendeten Spezies-Grenzwerten) wie auch nach außen (alle paarweisen Vergleiche eines beliebigen Stammes aus der Teilmenge mit einem beliebigen nicht zur Teilmenge gehörenden Stamm sind echt kleiner den angewendeten Spezies-Grenzwerten) auf Grundlage aller verwendeten Ähnlichkeitsmaße zur Speziesabgrenzung (hier ANI<sub>b1</sub>, ANI<sub>m</sub> und gANI) gibt. Zu beachten ist somit, dass die Wohl-Definiertheit einer Spezies immer abhängig vom betrachteten Datensatz, den verwendeten Indices zur Speziesabgrenzung und den angewendeten Spezies-Grenzwerten ist. Von den bisher bekannten Arten aus der *B. cereus s.l.* Gruppe zeichnen sich *B. "manliponensis"* (Block 1),





*B. cytotoxicus* (Block 4), *B. "gaemokensis"* (Block 5), *B. "bingmayongensis"* (Block 6), *B. pseudomycoides* (Block 9) und *B. toyonensis* (Block 19) als wohl-definierte Spezies aus, wobei die Arten *B. "bingmayongensis"*, *B. "gaemokensis"* und *B. "manliponensis"* lediglich durch ihre Typstämme (Singletons) repräsentiert werden. Daneben finden sich zehn weitere wohl-definierte und potentielle neue Genomospezies (Blöcke 2, 3, 7, 8, 10, 11, 13, 15, 17 und 18) im Datensatz. Die Mehrheit der neuen Arten (7 von 10) ist allerdings auch mit nur einem Stamm (Singleton) vertreten. Bei gesonderter Betrachtung jedes Index für sich rangiert die Anzahl potentieller neuer Genomospezies dahingegen zwischen  $\geq 10$  und  $\geq 17$  (Tabelle 21).

Die verbleibenden vier Blöcke 12, 14, 16 und 20 weisen Unterschiede in der Vorhersage der Spezies-Cluster zwischen den drei phylogénomischen Indices auf (grün hinterlegte Cluster). Unter Berücksichtigung dieser Unterschiede wurden die Blöcke in Teilcluster untergliedert, wobei jedes Teilcluster immer die Größe (Anzahl an Stämmen) des kleinsten vorhergesagten Spezies-Clusters eines Index einnimmt. Eine Ausnahme zu dieser Regel findet sich bei den beiden Teilclustern A und B von Block 20, in dessen monophyletischer Gruppe die beiden Typstämme der Arten *B. thuringiensis* und *B. cereus* s.s. anwesend sind. Hier fand die Eingruppierung der 106 Stämme anhand der höheren phylogénomischen Ähnlichkeit (aller drei Indices) entweder zum *B. thuringiensis* (Teilcluster 20-A) oder zum *B. cereus* s.s. (Teilcluster 20-B) Typstamm statt. Diese Teilcluster-Bildung wird auch durch die Topologie des MLSA ML-Baumes in eine *B. thuringiensis*-assoziierte Linie (Teilcluster 20-A) und eine *B. cereus* s.s.-assoziierte Linie (Teilcluster 20-B) mit eindeutigen Bootstrap-Werten (100 %) unterstützt. Hinsichtlich der Indices ANI<sub>b1</sub> und ANI<sub>m</sub> bilden die Stämme aus Block 20 ein Spezies-Cluster in Form einer Clique. Zudem weisen auch die beiden Typstämme im direkten Vergleich Ähnlichkeitswerte (ANI<sub>b1</sub> = 97,17 %, ANI<sub>m</sub> = 96,72 % und gANI = 96,94 %), die über den Spezies-Grenzwerten liegen, auf. Dies bekräftigt eine Reklassifizierung von *B. thuringiensis* als zugehörig zu *B. cereus* s.s. Da Block 20 allerdings keine wohl-definierte Spezies im Sinne dieser Arbeit ist (gANI sagt eine Cliquen-Gruppe vorher), wird eine Reklassifizierung von *B. thuringiensis* im Rahmen dieser Arbeit nicht vorgeschlagen. Vielmehr werden die 39 Stämme aus Teilcluster 20-A der Art *B. thuringiensis* zugeordnet, wohingegen die 67 Stämme aus Teilcluster 20-B als *B. cereus* s.s. Stämme angesehen werden.

Tabelle 20 Genomsequenzierungen und Assemblierungen von Stämmen der *B. cereus s.l.* Gruppe

ID	Stammbezeichner	Sequenzierte Read-Länge [bp]	Read Länge nach Qualitätskontrolle [bp]	Sequenziertiefe [x-fach]	Längstes Contig [bp]	N50	L50	Anzahl Contigs	Assembly Size
B001	<i>B. cytotoxicus</i> CVUAS 10162	2x251	2x220	23	579.447	187.878	7	72	4.126.479
B002	<i>B. cytotoxicus</i> CVUAS 3828	2x251	2x220	38	921.369	223.508	6	51	4.177.097
B003	<i>B. cytotoxicus</i> CVUAS 6082	2x251	2x220	46	1.194.431	288.705	4	56	4.143.745
B004	<i>B. cytotoxicus</i> CVUAS 644	2x251	2x220	35	1.242.095	292.929	4	66	4.220.310
B005	<i>B. cytotoxicus</i> CVUAS 6661	2x251	2x220	35	1.017.915	190.092	5	67	4.112.293
B006	<i>B. cytotoxicus</i> CVUAS 6688	2x251	2x220	24	618.717	153.575	7	86	4.213.628
B007	<i>B. mycooides</i> DSM 307	2x251	2x220	39	1.122.768	192.452	7	80	5.908.776
B008	<i>B. pseudomycooides</i> CVUAS 7789	2x251	2x220	39	788.204	228.066	8	100	5.776.115

Tabelle 21 Anzahl bekannter und potentieller neuer *B. cereus s.l.* Spezies

Ähnlichkeitsmaß	Gesamtanzahl an vorhergesagten			Anzahl bekannter Spezies in Form von			Anzahl potentieller neuer Spezies in Form von		
	Singletons	Cliquen	Cliquen-Gruppen	Singletons	Cliquen	Cliquen-Gruppen	Singletons	Cliquen	Cliquen-Gruppen
ANlb <sub>1</sub>	10	7 (158 Stämme)	3 (113 Stämme)	3 <sup>a</sup>	5 <sup>b</sup> (147 Stämme)	3 <sup>c</sup> (113 Stämme)	7	3 (11 Stämme)	≥ 0 <sup>d</sup>
ANlm	13	11 (199 Stämme)	1 (69 Stämme)	3 <sup>a</sup>	7 <sup>b,c</sup> (180 Stämme)	1 <sup>c</sup> (69 Stämme)	10	5 (19 Stämme)	≥ 0 <sup>d</sup>
gANI	13	11 (119 Stämme)	3 (149 Stämme)	3 <sup>a</sup>	5 <sup>c</sup> (95 Stämme)	3 <sup>b,c</sup> (116 Stämme)	10	6 (24 Stämme)	≥ 1 <sup>d</sup> (33 Stämme)

Statistik basierend auf der Annahme einer Reklassifizierung von *B. weihenstephanensis* zu *B. mycooides*.

<sup>a</sup>, *B. "bingmayongensis"*, *B. "gaemokensis"* und *B. "manliponensis"* wurden als bekannte Spezies mit berücksichtigt, obwohl die drei Arten (noch) nicht in die LPSN aufgenommen wurden.

<sup>b</sup>, *B. cereus s.s.* und *B. thuringiensis* wurden als zwei unterschiedliche Spezies gewertet.

<sup>c</sup>, *B. anthracis* und *B. wiedmannii* wurden auch als bekannte Arten mit einbezogen, obwohl die anerkannten Typstämme beider Spezies im *B. cereus s.l.* Datensatz nicht enthalten sind.

<sup>d</sup>, Aufgrund der fehlenden Auflösung von Cliquen-Gruppen in einzelne Spezies wurde die Anzahl neuer Arten in Cliquen-Gruppen mit mindestens einem Typstamm standardmäßig auf ≥ 0 festgesetzt. Bei Cliquen-Gruppen ohne Typstamm entsprechend auf ≥ 1.

Die monophyletische Gruppe von Block 12 beinhaltet die beiden anerkannten Typstämme *B. mycoides* ATCC 6462<sup>T</sup> und *B. weihenstephanensis* WSBC 10204<sup>T</sup>. Genauer gesagt sind beide Typstämme Mitglieder des Teilclusters 12-D, das von den beiden Indices ANIm und gANI als Clique vorhergesagt wurde. Im direkten Vergleich der Typstämme liegen die paarweisen Ähnlichkeitswerte aller drei Ähnlichkeitsmaße (ANIB<sub>11</sub> = 97,78 %, ANIm = 97,54% und gANI = 97,74 %) deutlich über den gesetzten Spezies-Grenzwerten aus Tabelle 7. Zudem ist auf Grundlage der Topologie des MLSA ML-Baumes eine klare Trennung der in Teilcluster 12-D befindlichen Stämme in eine *B. mycoides*-assoziierte Linie und eine *B. weihenstephanensis*-assoziierte Linie nicht möglich. Aus diesen Gründen (und wie bereits an anderer Stelle formal vorgeschlagen [413]) geht diese Arbeit von einer Reklassifizierung von *B. weihenstephanensis* als zugehörig zu *B. mycoides* aus. Ob die sieben Stämme aus den Teilclustern 12-A, 12-B und 12-C auch der Art *B. mycoides* hinzuzuzählen sind ist unwahrscheinlich. Gemäß ANIB<sub>11</sub> bilden zwar alle Stämme aus Block 12 eine Cliquen-Gruppen, allerdings grenzen sich die Teilcluster in den ANIm- und gANI-Vorhersagen klar als Clique (Teilcluster 12-C) bzw. Singletons (Teilcluster 12-A und 12-B) von Teilcluster 12-D ab.

Durch nachträgliche ANI-Screenings (Ergebnis-Abschnitt 3.2.1. ) konnte der *B. wiedmannii* Typstamm Teilcluster 14-C zugeordnet werden. Angesichts der Heterogenität der sich abbildenden Spezies-Custer in Block 14 gestaltet sich eine genaue Speziesabgrenzung allerdings schwierig. Relativ sicher erscheint aufgrund der ANIm- und gANI-Vorhersagen, dass die Mitglieder aus Teilcluster 14-A (Clique) sowie 14-B (Singleton) nicht mehr der Spezies *B. wiedmannii* zuzurechnen sind. Ob jedoch alle in Teilcluster 14-C befindlichen Stämme zur Art *B. wiedmannii* gezählt werden können, ist aufgrund der gANI-Vorhersage einer Cliquen-Gruppe ungewiss.

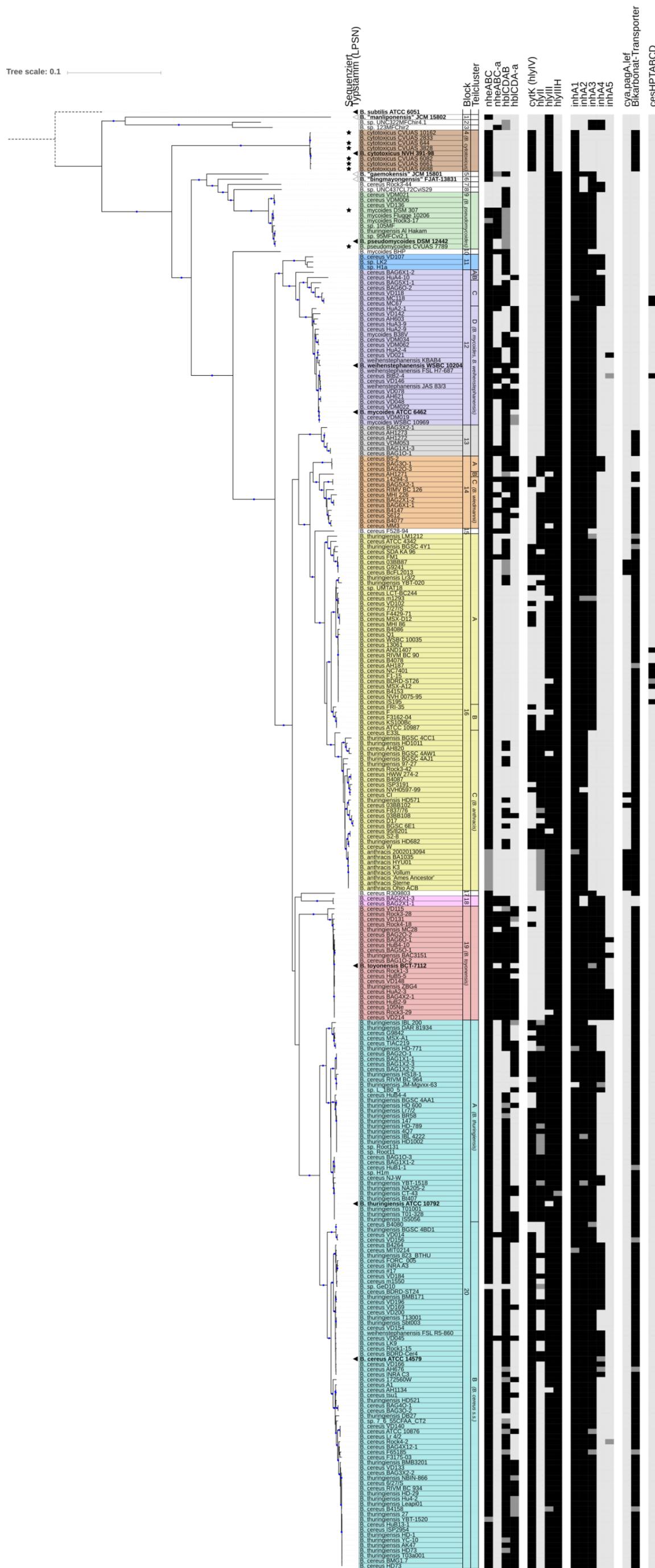
Der klonale *B. anthracis* Komplex bestehend aus den acht Stämmen *B. anthracis* 2002013094, *B. anthracis* 'Ames Ancestor', *B. anthracis* BA1035, *B. anthracis* HYU01, *B. anthracis* K3, *B. anthracis* Ohio ACB, *B. anthracis* Sterne und *B. anthracis* Vollum findet sich in Teilcluster 16-C wieder. Dieses Teilcluster wird durch den Index ANIg als Clique vorhergesagt, wohingegen bei den verbleibenden zwei Indices (ANIB<sub>11</sub>, ANIm) der gesamte Block 16 eine Cliquen-Gruppe bildet. Nach (künstlicher) Anhebung der Spezies-Grenzwerte auf 97,11 % (ANIB<sub>11</sub>) bzw. 96,91 % (ANIm) zeichnet sich Teilcluster 16-C jedoch auch in den beiden anderen Fällen als Clique aus. Damit ist allen ANI-Vorhersagen gemeinsam, dass die Art *B. anthracis* sich über

den klonalen Komplex hinaus auf zumindest die in Teilcluster 16-C ansässigen Stämme erstreckt.

### **3.3. Verteilung und Eignung von *B. cereus s.l.* Virulenzfaktoren als Spezies-diskriminierende Marker**

#### **3.3.1. Untersuchte Virulenzfaktoren**

Die untersuchten Virulenzfaktoren aus der *B. cereus s.l.* Gruppe wurden in unterschiedliche Klassen eingeteilt. Zu diesen Klassen gehören die Enterotoxin-Komplexe (Nhe, HBL), die Hämolsine (CytK, HlyII, HlyIII, HlyA), die Metalloproteasen (InhA1, InhA2, InhA3), das Anthrax-Toxin (EF, PA, LF) zuzüglich des Bikarbonat-Transporters, sowie die Proteine des Cereulid-Synthetase-Genclusters (*cesHPTABCD*). Neben diesen für eine Toxikoinfektion bzw. Intoxikation des Menschen bedeutsamen Virulenzfaktoren wurden auch die Bt-Toxine (Cry, Cyt, Vip, Sip) mit ihrer insektiziden Wirkung als eigene Klasse untersucht. Während der Suche konnte jedem Virulenzfaktor mindestens eine vorhergesagte homologe Proteinfamilie zugeordnet werden (siehe Methoden-Abschnitt 14.3. ). Die mit den einzelnen Virulenzfaktoren assoziierten Proteinfamilien wurden schließlich im Detail analysiert. So konnte für jeden Stamm des *B. cereus s.l.* Datensatzes ein An-/Abwesenheitsprofil der genetischen Information (CDS) über alle Virulenzfaktoren hinweg erstellt werden. Die entsprechenden An-/Abwesenheitsprofile (ohne Bt-Toxine) sind in Abbildung 14 dargestellt. Dahingegen sind die Ergebnisse der Suche nach Bt-Toxinen gesondert in Abbildung 16 und Abbildung 17 als Zählung von Cry, Cyt und Vip Toxinen pro Stamm zu finden. Die nachfolgenden Abschnitte fassen die wichtigsten An-/Abwesenheits-Charakteristika und -Tendenzen, die für jede Virulenzfaktor-Klasse gemacht werden konnten, zusammen. Als Ergänzung zu dieser Klassen-basierten Abhandlung dient der letzte Abschnitt des Kapitels. Dort werden die gemachten Erkenntnisse für jede der bekannten *B. cereus s.l.* Spezies mit bereits erfolgter Aufnahme in die LPSN in Tabelle 23 (Seite 126) zusammengefasst. Dies soll den Leser zu einer schnellen Inaugenscheinnahme der Charakteristika und Tendenzen, die sich im Zusammenhang mit der Verteilung der Virulenzfaktoren bei diesen Arten ergeben, befähigen.



**Abbildung 14 Verteilung von *B. cereus s.l.* Virulenzfaktoren**

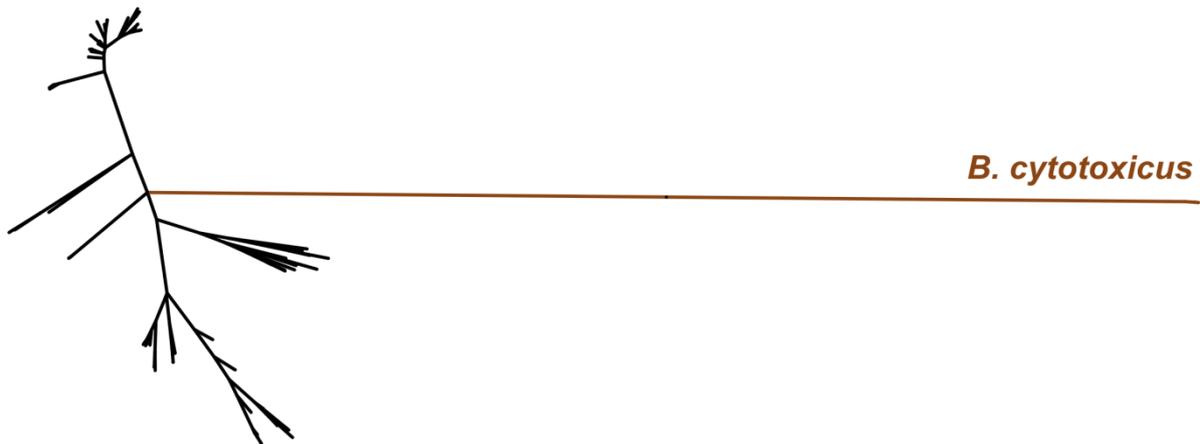
Dargestellt ist die MLSA Phylogenie des gesamten *B. cereus s.l.* Datensatzes (links). Der Phylogenie liegen die kompletten CDS der Gene *adh*, *ccpA*, *glpF*, *panC*, *pta* und *pycA* zugrunde. Die Wurzelung des Baumes fand mit der Außengruppe *B. subtilis* ATCC 6051<sup>T</sup> statt. Zur übersichtlicheren Darstellung der Topologie wurden die phylogenetischen Distanzen zur Außengruppe künstlich auf 0,1 Substitutionen pro Alignmentposition gesetzt (gestrichelte Äste). Unter Annahme des GTR+G+I Modells wurde mittels der ML-Methode die Phylogenie mit 500 Bootstrap-Replikaten rekonstruiert. Äste mit hoher Bootstrap-Unterstützung ( $\geq 70\%$ ) sind mit blauen Kreisen gekennzeichnet. Eigens *de novo* sequenzierte und assemblierte Genome weisen einen schwarzen Stern auf. Schwarze Dreiecke stehen für Typstämme valide beschriebener Arten (LPSN Stand November 2017). Typstämme von Spezies ohne bisherige Aufnahme in die LPSN weisen dahingegen leere Dreiecke auf. Alle Stämme wurden gemäß ihrer Blockzugehörigkeit farblich gleichartig hinterlegt. Weiß gehalten sind Stämme von Blöcken, die nur aus einem einzigen Stamm bestehen. Sofern bestimmte Blöcke noch in Teilcluster untergliedert wurden, ist auch dies gekennzeichnet (Spalte Teilcluster). Die An-/Abwesenheitsprofile der Virulenzfaktor-CDS (auf Nucleinsäureebene) der einzelnen Stämme finden sich rechts in der Graphik. Sie sind gemäß ihrer Klasse in der Reihenfolge Enterotoxin-Komplexe (*nheABC*, *nheABCa*, *hblCDAB*, *hblCDAa*), Hämolsine (*cytK*, *hlyII*, *hlyIII*, *hlyVIII*), Metalloproteasen (*inhA1*, *inhA2*, *inhA3*, *inhA4*, *inhA5*), Anthrax-Toxin (*cya*, *pagA*, *lef*) inklusive Bikarbonat-Transporter und Cereulid-Synthetase-Gencluster (*cesHPTABCD*) sortiert. Schwarze Zellen in den Profilen stehen für die Anwesenheit des Genes des jeweiligen Faktors über die gesamte Länge. Bei dunkelgrauen Zellen liegt das Gen entweder nur teilweise im Genom des entsprechenden Stammes vor (z.B. aufgrund vorzeitiger Beendigung des kodierenden Contigs, oder schlechter Qualität des kodierenden Contigs), oder der Stamm weist Mutationen, die zu einem unvollständigen Genprodukt führen (z.B. Nonsense- oder Frameshift-Mutationen), auf. Sofern dem betrachteten Virulenzfaktor ein Gencluster/Operon zugrunde liegt, reicht die Unvollständigkeit bzw. Mutation eines einzelnen Genes aus, um das gesamte Gencluster/Operon dunkelgrau zu markieren. Hellgraue Zellen dagegen stehen für die gänzliche Abwesenheit der CDS des jeweiligen Faktors.

### 3.3.2. Das Anthrax-Toxin und die Art *B. anthracis*

Neben den acht Stämmen des klonalen *B. anthracis* Komplexes enthalten fünf weitere Stämme des Datensatzes die Gene für das Anthrax-Toxin (*cya*, *pagA*, *lef*). Zwei der Stämme (*B. cereus* CI und *B. cereus* 03BB102) sind wie der klonale Komplex ansässig in Teilcluster 16-C (Abbildung 14). Die verbleibenden drei Stämme (*B. cereus* 03BB87, *B. cereus* G9241 und *B. cereus* BcFL2013) kommen dahingegen aus Teilcluster 16-A. Die Präsenz der Anthrax-Toxin Gene kann somit nicht als Alleinstellungsmerkmal des klonalen *B. anthracis* Komplexes angesehen werden. Da zudem nicht alle Stämme aus Teilcluster 16-A über die genetische Information für das Anthrax-Toxin verfügen, kann auch die Genomospezies *B. anthracis* nicht allein über das Anthrax-Toxin definiert werden. Damit weist die alleinige Präsenz der Anthrax-Gene in einem Stamm diesen nicht zwingendermaßen als Mitglied des klonalen Komplexes aus. Noch ist die Abwesenheit der Gene ein ausschlaggebendes Merkmal dafür, dass es sich nicht um ein Isolat der Genomospezies *B. anthracis* handeln kann.

Es ist bekannt, dass die Anwesenheit von Bikarbonat einen positiven Effekt auf die Transkription der Anthrax-Gene (*cya*, *pagA*, *lef*) bei Stämmen des klonalen *B. anthracis* Komplexes hat [240, 414]. Deshalb wurde auch die Verteilung des Bikarbonat-Transporters im *B. cereus* s.l. Datensatz untersucht (Abbildung 14). Der Transporter setzt sich aus drei unterschiedlichen Proteinen (ATP-Bindeprotein, Permease, Substrat-Bindeprotein) zusammen, deren CDS in einem Gencluster festgeschrieben sind. Jede der drei Proteinkomponenten ist Teil einer eigenen homologen Familie. Nach Analyse der drei Familien war aufgefallen, dass die Art *B. cytotoxicus* (Block 4) über eine eigene Variante des Transporters verfügt (Abbildung 15). Weiterhin liegt der Transporter in jedem Stamm der Blöcke 15, 16, 17, 19 und 20 vor und ist somit in den Spezies *B. anthracis* (Teilcluster 16-C), *B. toyonensis* (Block 19), *B. thuringiensis* (Teilcluster 20-A) und *B. cereus* s.s. (Teilcluster 20-B) strikt anwesend. Demgegenüber fehlt der Transporter in einem Stamm (*B. cereus* BAG3X2-1) aus Block 13 sowie vier Stämmen (*B. cereus* B5-2, *B. cereus* BAG3O-1, *B. cereus* BAG2O-3 und *B. cereus* BAG5X2-1) aus Block 14. In den verbleibenden Blöcken und Teilclustern (darunter die Arten *B. mycoides* und *B. pseudomycoides*) ist der Transporter auch nicht vorhanden.

Tree scale: 0.1



**Abbildung 15 Phylogenie des Substrat-Bindeprotein Genes des Bikarbonat-Transporters**

Die Abbildung zeigt die ungewurzelte ML-Phylogenie basierend auf dem Gen für das Substrat-Bindeprotein des Bikarbonat-Transporters. Die phylogenetische Rekonstruktion fand unter Annahme des GTR+G+I Modells statt. Unvollständige Gensequenzen sind in der Phylogenie nicht berücksichtigt. Die Art *B. cytotoxicus* besitzt eine eigene Form des Genes (braun). Dies ist auch bei der Permease und dem ATP-Bindeprotein des Transporters der Fall (nicht gezeigt).

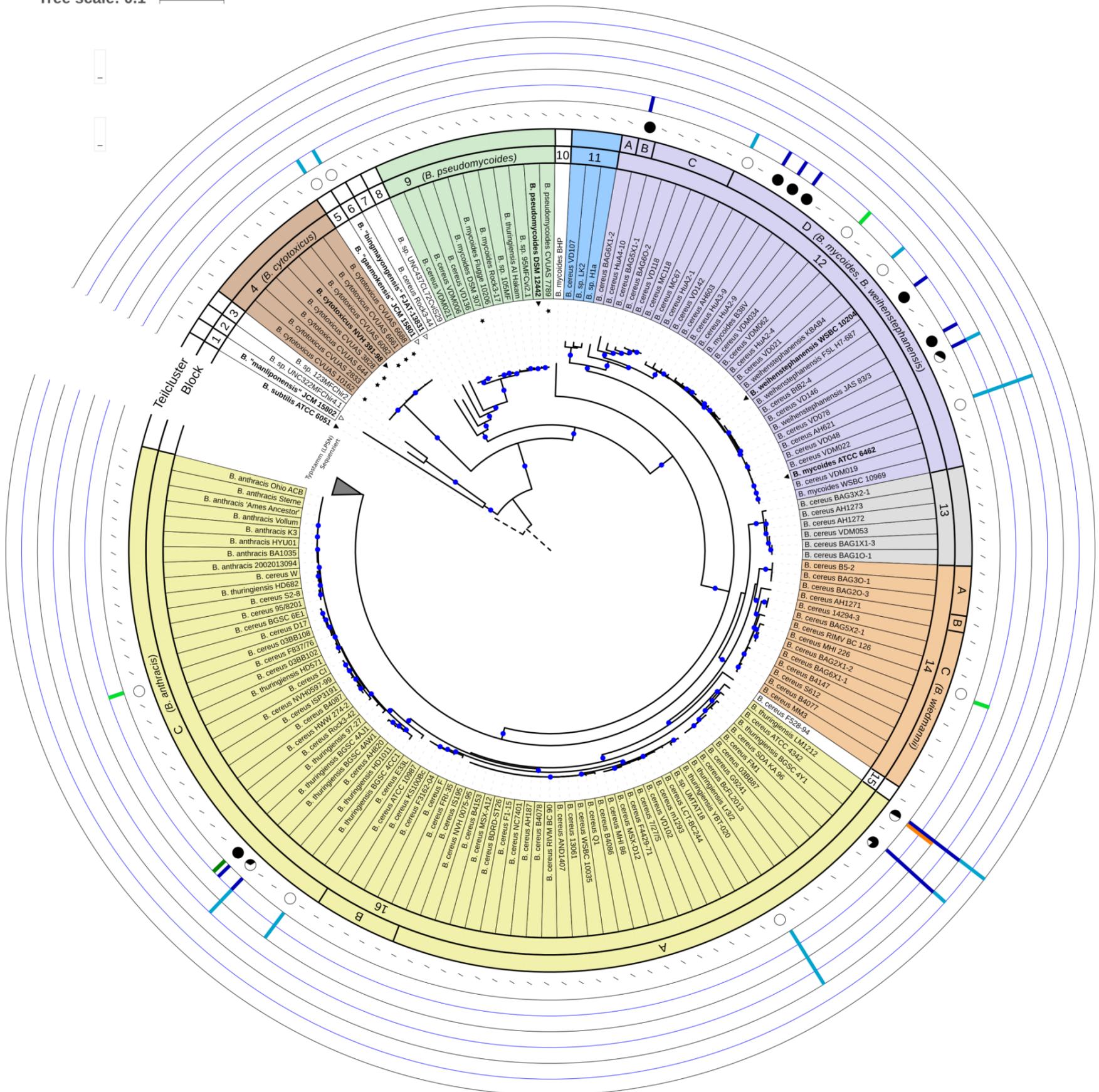
### 3.3.3. *B. thuringiensis* und seine Bt-Toxine

Von den 348 bekannten Holotyp-Sequenzen (Stand Februar 2016) konnten 332 (95,4 %) vom NCBI bezogen und für die Suche im *B. cereus* s.l. Datensatz verwendet werden. Dabei waren die Holotyp-Abdeckungsgrade der einzelnen Toxin-Klassen durchgehend hoch (Cry: 96,05 %, Cyt: 100 %, Vip: 87,5 %, Sip: 100 %). Neben den bekannten Toxin-Holotypen konnten im Rahmen dieser Arbeit 22 potentielle neue Holotypen ausfindig gemacht werden.

Abbildung 16 und Abbildung 17 zeigen die Verteilung der Cry, Cyt und Vip Toxine im Datensatz. Das bis dato einzig bekannte Sip Toxin (SipAa1) konnte in nur in einem Stamm (*B. cereus* TIAC219 aus Teilcluster 20-A) gefunden werden. Die Art *B. thuringiensis* besitzt mit ~ 54 % zwar die höchste relative Anzahl an Stämmen mit mindestens einem Bt-Toxin, dennoch ist das Vorkommen von Bt-Toxin kodierenden Stämmen nicht auf Teilcluster 20-A beschränkt. So ist der Anteil an Stämmen mit mindestens einem Bt-Toxin bei den beiden Arten *B. cereus* s.s (~ 37 % in Teilcluster 20-B) und *B. mycoides* (~ 26 % in Teilcluster 12-D) nicht unerheblich. Zudem weisen zwölf Stämme (~ 31 %) aus Teilcluster 20-A nicht einmal die genetische Information für ein putatives Bt-Toxin auf. Sowohl die Tatsache, dass nicht alle Stämme aus Teilcluster 20-A über ein Bt-Toxin verfügen, als auch die gehäufte und unzweifelhafte Präsenz von Bt-Toxinen in Isolaten anderer *B. cereus* s.l. Spezies, decken eine

Verwendung von Cry, Cyt, Vip und Sip Proteinen als Spezies-definierendes Merkmal für die Art *B. thuringiensis* als ungeeignet auf. Dieser Rückschluss bleibt selbst bei Betrachtung von *B. thuringiensis* und *B. cereus* s.s. als ein und dieselbe Spezies im Grundsatz erhalten.

Tree scale: 0.1



**Abbildung 16 Verteilung von Bt-Toxinen (Teilbaum 1)**

Die gezeigte Graphik ist bis zum fünften Ring gleichbedeutend mit Abbildung 12. Im sechsten Ring ist für jeden Stamm das Verhältnis der Anzahl gefundener Toxine zu putativen Toxinen dargestellt. Dementsprechend besitzen Stämme mit einem schwarzen Kreis ausschließlich Bt-Toxine, deren Sequenzähnlichkeit  $\geq 95\%$  (auf Aminosäureebene) zu einem bereits bekannten Toxin-Holotyp ist. Weiße Kreise dahingegen stehen für Stämme in denen einzig putative Bt-Toxine gefunden werden konnten. Ein Protein gilt als putatives Bt-Toxin, sofern es verkürzt vorliegt (Nonsense-Mutation, Frameshift-Mutation, Ende des kodierenden Contigs, etc.), oder Mitglied einer Bt-Toxin-assoziierten homologen Familie mit weniger als  $95\%$  Sequenzähnlichkeit zum ähnlichsten bekannten Holotyp ist. Der skizzierten Regel folgend weisen Stämme mit einer gleichgroßen Anzahl an Toxinen und putativen Toxinen einen schwarzen Halbkreis auf. Die Anzahl gefundener Cry, Cyt oder Vip Toxine pro Stamm ist im sechsten Ring als Säulendiagramm aufgeführt (Skalierung von 1 bis 6 in 1er-Schritten). Dabei werden die drei Toxin-Arten durch farblich unterschiedliche Balken repräsentiert. Dunkelfarbige Balken stehen hierbei für Toxine (Cry: blau, Cyt: rot, Vip: grün) und hellfarbige Balken für putative Toxine (Cry putativ: hellblau, Cyt putativ: orange, Vip putativ: hellgrün). Kommen sowohl Toxine als auch putative Toxine innerhalb eines Stammes vor, sind die Balken der betroffenen Toxin-Art gestapelt.



### 3.3.4. Die Enterotoxin-Komplexe Nhe und HBL

Die beiden Enterotoxin-Komplexe Nhe und HBL zählen zu den Hauptvirulenzfaktoren bei durch *B. cereus* verursachten Durchfallerkrankungen. Bei beiden Enterotoxin-Komplexen wurde die Existenz homologer Komplexe ( $Nhe_a$ ,  $HBL_a$ ) bereits nachgewiesen [26, 306]. Die Gene der Enterotoxin-Komplexe liegen als Operons (*nheABC*, *hblCDAB*) im Genom vor. Dies ist auch bei den homologen Komplexen der Fall, wobei dem *hblCDA<sub>a</sub>* Operon das *hblB* Gen fehlt.

Die Häufigkeit des *nheABC* Operons ist mit ~ 97 % unter den Stämmen des *B. cereus s.l.* Datensatzes sehr hoch (Abbildung 14). Lediglich acht Stämme besitzen das Operon nicht, darunter eine Klade (*B. cereus* VDM021, *B. cereus* VDM006 und *B. cereus* VD136) aus Block 9 (*B. pseudomycooides*). Zudem tragen alle Stämme des klonalen *B. anthracis* Komplexes eine Nonsense-Mutation im Gen *nheC*. Im Gegensatz zur weiten Verbreitung von *nheABC* kommt das Gencluster des homologen Komplexes *nheABC<sub>a</sub>* in weit weniger Stämmen vor. Dieses weisen nur ~ 18,5 % aller Stämme auf. Hier ist besonders auffällig, dass ~ 86 % der *B. toyonensis* Stämme (Block 19) über *nheABC<sub>a</sub>* verfügen. In den Blöcken 9 (*B. pseudomycooides*) und 12 hat annähernd jeder zweite Stamm das homologe Operon, wohingegen es nur drei der 175 Stämme aus den Blöcken 16 und 20 (*B. thuringiensis* und *B. cereus s.s.*) vorweisen. In der Art *B. cytotoxicus* (Block 4) ist *nheABC<sub>a</sub>* gänzlich abwesend.

Die Häufigkeit von *hblCDAB* liegt bei ~ 70,5 % für den gesamten Datensatz. Dahingegen kommt das homologe Operon mit ~ 33,5 % annähernd nur halb so oft vor. Wie schon im Falle von Nhe ist der Anteil an Stämmen mit beiden HBL Operons bei der Art *B. toyonensis* (Block 19) beachtlich. So tragen alle *B. toyonensis* Stämme *hblCDAB* und ~ 86 % zudem *hblCDA<sub>a</sub>*. Dem folgt Block 14 mit überdurchschnittlichen Häufigkeiten von ~ 93 % (*hblCDAB*) bzw. ~ 71% (*hblCDA<sub>a</sub>*). In Block 12 besitzen zwei von drei Stämmen die genetische Information für HBL. Das gleiche gilt auch für den homologen Komplex. Die beiden Spezies *B. thuringiensis* und *B. cereus s.s.* (Block 20) weisen einen hohen Anteil an *hblCDAB* kodierenden Stämmen auf (~ 79,5 % bzw. ~ 97 %). Im Vergleich dazu tragen bei beiden Spezies weit weniger Stämme das homologe Gencluster (~ 36 % bzw. ~ 45 %). In Block 16 können lediglich 19 der 69 Stämme (~ 27,5 %) die Gene für HBL vorweisen. Zudem besitzen die drei Anthrax-Toxin kodierenden Stämme *B. cereus* 03BB87, *B. cereus* G9241 und *B. cereus* BcFL2013 aus Teilcluster 16-A eine Punktmutation im Start-Codon (ATG → ATA) des *hblD* Genes. Lediglich in einem einzigen Stamm (*B. cereus* 03BB108) aus

Teilcluster 16-C (*B. anthracis*) konnte auch das *hblCDA<sub>a</sub>* Operon gefunden werden. Der Art *B. cytotoxicus* (Block 4) fehlen beide HBL Ausprägungen. Im Gegensatz dazu konnten in den Stämmen der Spezies *B. pseudomycooides* (Block 9) zumindest Teile des *hblCDAB* Operons gefunden werden. Allerdings ist auf Grundlage der vorliegenden Genom Assemblierungen stark davon auszugehen, dass die Stämme *B. mycooides* DSM 307, *B. mycooides* Flugge 10206, *B. mycooides* Rock3-17, *B. sp.* 95MFCvi2.1, *B. pseudomycooides* DSM 12442<sup>T</sup> und *B. pseudomycooides* CVUAS 7789 nicht zur Bildung eines funktionsfähigen Komplexes im Stande sind.

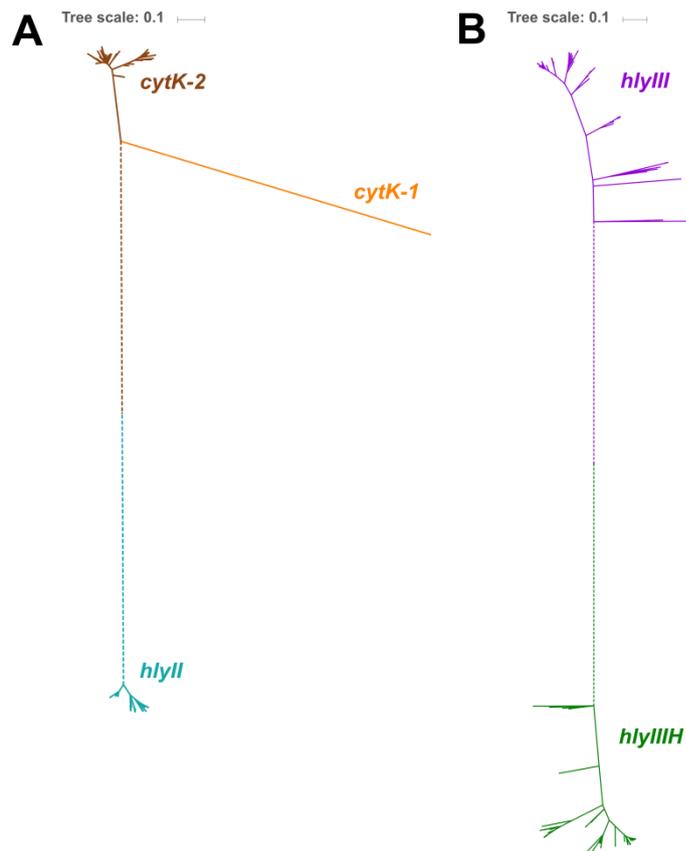
### 3.3.5. Die Hämolyse CytK, HlyII, HlyIII und HlyA

Neben den beiden Komplexen Nhe und HBL gilt auch das Hämolyse CytK als eines der wichtigsten Enterotoxine bei durch *B. cereus* induzierten Lebensmittelvergiftungen mit Durchfallerscheinungen. Die Art *B. cytotoxicus* verfügt über eine eigene Variante des Toxins, die mit CytK-1 bezeichnet wird [167, 208, 312].

Die beiden Hämolyse CytK und HlyII konnten einer vorhergesagten homologen Familie zugeordnet werden. Die Sequenzähnlichkeit der beiden Proteine liegt bei ~ 37 % (CytK-1 vs. HlyII) bzw. ~ 39 % (CytK-2 vs. HlyII). Abbildung 18A zeigt die ML-Phylogenie basierend auf den Gensequenzen beider Toxine im *B. cereus s.l.* Datensatz. Hinsichtlich der Verbreitung von *cytK* ist eine Zweiteilung der Häufigkeiten (hoch oder niedrig) bei den valide beschriebenen Arten zu beobachten (Abbildung 14). So kommt das Gen in *B. pseudomycooides* (Block 9) und *B. mycooides* (Teilcluster 12-D) überhaupt nicht vor. In *Bacillus toyonensis* (Block 19) und *B. wiedmannii* (Teilcluster 14-C) besitzen nur ~ 9 % bzw. ~ 20 % die CDS. In starkem Gegensatz dazu stehen die Arten *B. thuringiensis* (Teilcluster 20-A), *B. cereus* s.s. (Teilcluster 20-B) und *B. cytotoxicus* (Block 4) bei denen der Anteil *cytK* kodierender Stämme zwischen ~ 85 % und 100 % rangiert. Einzig *B. anthracis* (Teilcluster 16-C) liegt mit ~ 48 % seiner Stämme im Bereich der durchschnittlichen Häufigkeit von *cytK* (~ 50 %), die sich unter Berücksichtigung aller Stämme für den *B. cereus s.l.* Datensatz ergibt.

Wie im Falle von *cytK* liegt auch bei *hlyII* eine bimodale Verteilung vor (Abbildung 14). Strikt abwesend ist das Gen in den Spezies *B. cytotoxicus* (Block 4), *B. pseudomycooides* (Block 9), *B. mycooides* (Teilcluster 12-D) und *B. toyonensis* (Block 19). Demgegenüber weisen alle *B. thuringiensis* Isolate (Teilcluster 20-A) *hlyII* auf. Auffallend ist, dass das Gen in der zweiten Spezies (*B. cereus* s.s.) aus demselben Block in nur ~ 13 % der Stämme vorliegt. Ein ähnliches Bild ergibt sich auch für Block

16. Hier besitzen lediglich acht von 38 Isolaten aus den Teilclustern 16-A und 16-B die genetische Information für *hlyII*. Bei der Art *B. anthracis* (Teilcluster 16-C) fehlt das Gen jedoch in nur einem (*B. cereus* D17) von 31 Stämmen. Des Weiteren zeichnet sich der klonale *B. anthracis* Komplex durch eine Frameshift-Mutation, die ein trunkiertes HlyII Protein zur Folge haben sollte, aus. Auch die Art *B. wiedmannii* (Teilcluster 14-C) gehört zu den Vertretern mit einem hohen Vorkommen (70 %) an *hlyII* tragenden Stämmen.



**Abbildung 18** Phylogenien basierend auf den Genen *cytK* und *hlyII* bzw. *hlyIII* und *hlyIIIIH*

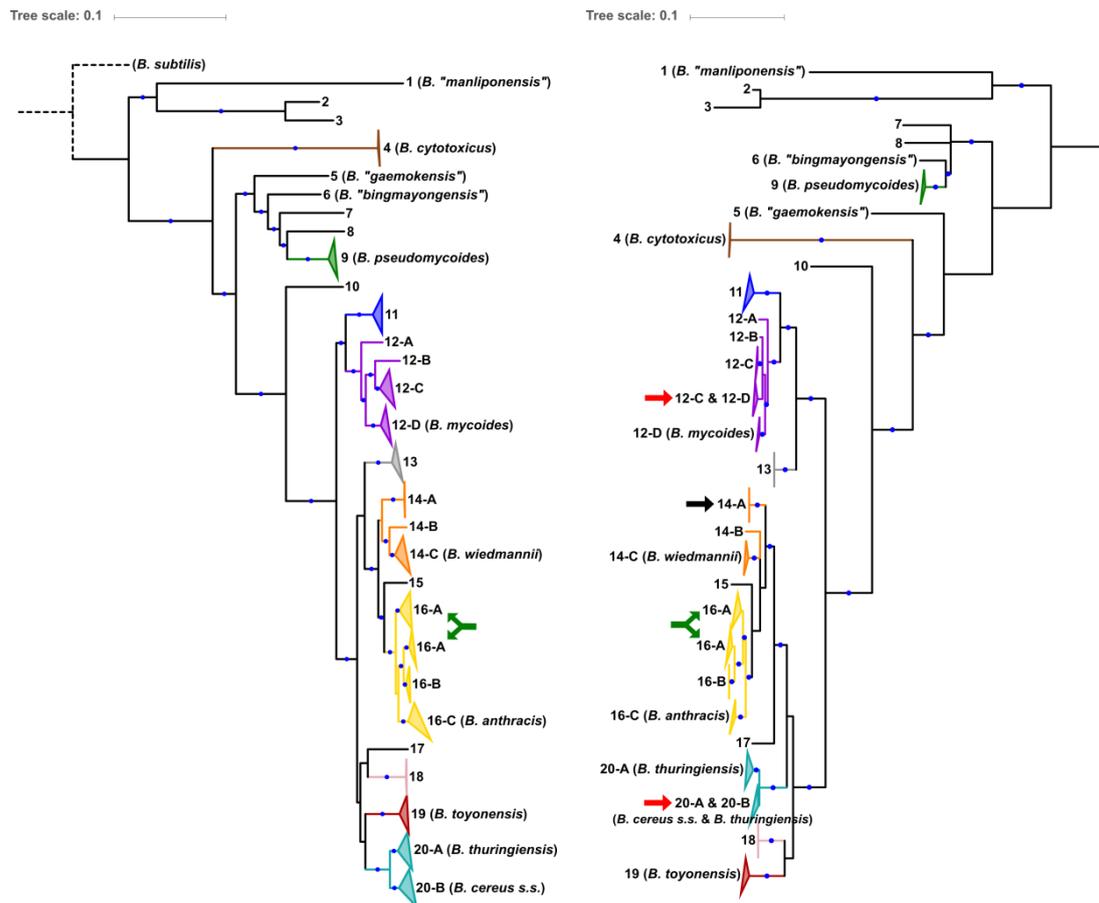
Beide Phylogenien wurden mit Hilfe der ML-Methode unter Annahme des GTR+G+I Substitutionsmodells erstellt. **A** Ungewurzelte Phylogenie basierend auf den Genen *cytK* (braun bzw. orange) und *hlyII* (dunkles türkis). Partielle Gensequenzen wurden von vornherein ausgeschlossen. Die phylogenetischen Distanzen gestrichelter Äste wurden künstlich von 176,44 auf 0,5 Substitutionen pro Alignmentposition reduziert. Beide Hämolsyne CytK und HlyII sind Teil einer vorhergesagten homologen Familie. **B** Ungewurzelte Phylogenie basierend auf den Gensequenzen der vorhergesagten homologen Familie in der HlyIII ansässig ist (magenta: *hlyIII*, grün: *hlyIIIIH*). Unvollständige Gensequenzen wurden nicht berücksichtigt. Die phylogenetischen Distanzen gestrichelter Äste wurden künstlich von 8,82 auf 0,5 Substitutionen pro Alignmentposition gesetzt. In Analogie zu CytK und HlyII umfasst die homologe Familie wahrscheinlich zwei unterschiedliche Hämolsyne. Das bis dato unbekanntes Hämolsyn wird als Hämolsyn III Homolog (HlyIIIIH) bezeichnet.

HlyIII ist Mitglied einer vorhergesagten homologen Familie in der noch ein zweites bisher nicht annotiertes Protein auftritt (Abbildung 18B). Dieses wurde mit der

Bezeichnung Hämolysin III Homolog (HlyIIIH) versehen. Die Sequenzähnlichkeit von HlyIII zu HlyIIIH beträgt ~ 42 %. Ob HlyIIIH tatsächlich hämolytisch aktiv ist und bei *B. cereus* Toxikoinfektionen mitwirkt, konnte im Rahmen dieser Arbeit nicht aufgeklärt werden. Die Klassifizierung des Proteins als Virulenzfaktor ist damit rein hypothetischer Natur und ergibt sich einzig aus der Ähnlichkeit, die die Phylogenie von *hlyIII* und *hlyIIIH* zur Phylogenie von *cytK* und *hlyII* aufweist (vergleiche Abbildung 18B mit Abbildung 18A). Die Gene *hlyIII* und *hlyIIIH* sind im Großteil der Stämme des *B. cereus s.l.* Datensatzes vorhanden (jeweils ~ 96 %). Stämme ohne *hlyIII* treten nur vereinzelt und verstreut über mehrere Blöcke und Teilcluster auf (Abbildung 14). In starkem Kontrast dazu fehlt *hlyIIIH* ausschließlich in den Blöcken 1 bis 5, während es in allen Stämmen der restlichen Blöcke anwesend ist.

In der vorhergesagten Proteinfamilie in der HlyA beheimatet ist, kommen keine paraloge Proteine vor. Jeder Stamm des *B. cereus s.l.* Datensatzes trägt exakt eine Kopie des *hlyA* Genes im Genom. Diese Beobachtung legt nahe, dass die Gensequenzen in den unterschiedlichen Spezies der *B. cereus s.l.* Gruppe in einer eins-zu-eins orthologen Beziehung zueinander stehen. Eins-zu-eins Orthologie wiederum ist ein Hinweis darauf, dass die Evolution von *hlyA* einer vertikalen Vererbungslinie folgt. Diese Eigenschaft könnte *hlyA* zu einem geeigneten Spezies Marker auf Sequenzebene machen. Abbildung 19 zeigt eine Gegenüberstellung der MLSA Phylogenie zur Einzelgen Phylogenie basierend auf *hlyA*. In der Einzelgen Phylogenie bilden sich zum größten Teil die gleichen Blöcke und Teilcluster, die sich auf Grundlage der ANI Speziesvorhersagen für die MLSA Phylogenie ergeben, ab. Unterschiede zwischen den Bäumen finden sich allein bei den Blöcken 12, 14 und 20. So formen jeweils zwei Stämme aus Teilcluster 12-C und 12-D eine Klade, womit die vier Stämme im *hlyA* Genbaum eine geringere phylogenetische Distanz untereinander aufweisen als zu den restlichen Isolaten gleicher Teilclusterzugehörigkeit (roter Pfeil). Die Stämme aus Block 14 stellen aufgrund der Abspaltung von Teilcluster 14-A keine monophyletische Gruppe mehr dar (schwarzer Pfeil) und es tritt eine teilweise Durchmischung von *B. thuringiensis* (Teilcluster 20-A) und *B. cereus s.s.* (Teilcluster 20-B) Isolaten auf (roter Pfeil). Vor allem diese letzte Beobachtung würde sich nachteilig auf eine Verwendung von *hlyA* als Spezies Marker auswirken, sofern *B. thuringiensis* und *B. cereus s.s.* tatsächlich als zwei eigenständige Arten angesehen werden (siehe Ergebnis-Abschnitt 3.2.2. ). Daneben ist auch die Aufspaltung von Teilcluster 16-C in zwei voneinander getrennte Cluster unvorteilhaft (grüner Doppelpfeil). Allerdings konnte auch schon auf Basis der MLSA

Phylogenie eine klare Trennung der drei Teilcluster aus Block 16 nicht erreicht werden.



**Abbildung 19** Gegenüberstellung der *B. cereus s.l.* MLSA und *hlyA* ML-Phylogenien

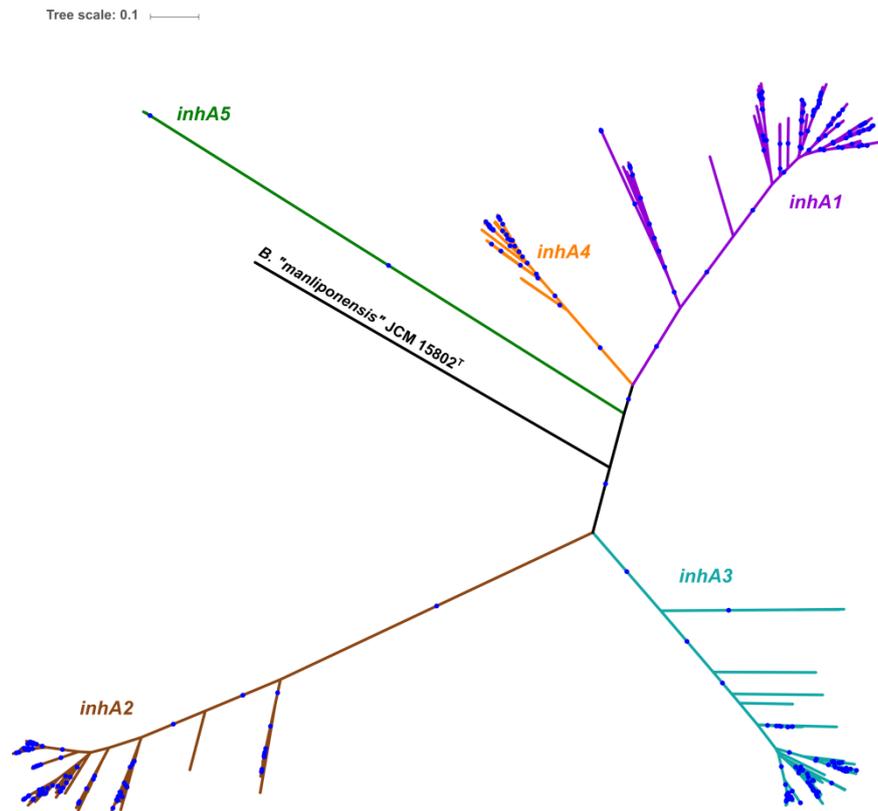
MLSA Phylogenie (links) und *hlyA* Einzelgen Phylogenie (rechts) für den *B. cereus s.l.* Datensatz. Unter Annahme des GTR+G+I Substitutionsmodells wurden beide Phylogenien mit der ML-Methode berechnet. Blaue Kreise kennzeichnen Äste mit hoher Bootstrap-Unterstützung ( $\geq 70\%$  von 500 Replikaten). Zur Wurzelung des MLSA Baumes diente die Außengruppe *B. subtilis* ATCC 6051<sup>T</sup>, wohingegen beim *hlyA* Genbaum eine Mittelpunktbewurzelung stattfand. Die phylogenetischen Distanzen zur Außengruppe des MLSA Baumes wurden künstlich auf 0,1 Substitutionen pro Alignmentposition reduziert. In beiden Bäumen sind die Blöcke und Teilcluster, die auf Grundlage der ANI Speziesvorhersagen definiert wurden, als zusammengefasste Knoten dargestellt. Grundsätzlich grenzen sich die Blöcke/Teilcluster in beiden Bäumen gut voneinander ab. Einzige Ausnahme bildet Teilcluster 16-A, das sich in beiden Bäumen in zwei Cluster aufspaltet (grüner Doppelpfeil). Daneben gibt es lediglich drei Unterschiede zwischen den Phylogenien. So tritt eine Durchmischung von Stämmen aus Teilcluster 12-C und 12-D als auch von Stämmen aus Teilcluster 20-A (*B. thuringiensis*) und 20-B (*B. cereus s.s.*) im *hlyA* Genbaum auf (rote Pfeile). Zudem bilden die Isolate aus Block 14 keine monophyletische Gruppe mehr (schwarzer Pfeil). Dies kann allerdings vernachlässigt werden, da Teilcluster 14-A als abgegrenzte und strukturelle Einheit im *hlyA* Genbaum erhalten bleibt.

### 3.3.6. Die Metalloproteasen InhA1, InhA2 und InhA3

Die Metalloproteasen InhA1, InhA2 und InhA3 sind Mitglieder der gleichen vorhergesagten homologen Proteinfamilie. Die Gene der Immuninhibitoren sind im *B.*

*cereus s.l.* Datensatz weit verbreitet. So tragen ~ 82 % der Stämme alle drei Gene und es gibt keinen Stamm, der nicht zumindest über die genetische Information für eine Metalloprotease verfügt. Zu den valide beschriebenen Arten, die *inhA1*, *inhA2* und *inhA3* besitzen, gehören *B. pseudomycooides* (Block 9), *B. mycooides* (Teilcluster 12-D), *B. wiedmannii* (Teilcluster 14-C), *B. toyonensis* (Block 19) und *B. cereus s.s.* (Teilcluster 20-B). Bei *B. thuringiensis* (Block 20-A) treten lediglich zwei Stämme auf, die *inhA1*- (*B. thuringiensis* JM-Mgvxx-63) bzw. *inhA3*-defizient (*B. thuringiensis* IBL 200) sind, ansonsten weisen auch hier sämtliche Isolate die drei CDS auf. Einzig *B. cytotoxicus* (Block 4) verfügt über ein relativ spezifisches An-/Abwesenheitsprofil, das sich mit *inhA1*-positiv, sowie *inhA2*- und *inhA3*-negativ beschreiben lässt. Lediglich ein Stamm (*B. cereus* NVH0597-99) aus Teilcluster 16-C weist das gleiche Präsenz/Absenz-Muster auf. Die restlichen *B. anthracis* Stämme (Teilcluster 16-C) sind allesamt *inhA3*-defizient. Das zugrundeliegende Profil (*inhA1*-positiv, *inhA2*-positiv, *inhA3*-negativ) stellt allerdings kein Alleinstellungsmerkmal der Art *B. anthracis* dar, da es mit drei weiteren Stämmen (*B. mycooides* BHP, *B. cereus* F528-94 und *B. thuringiensis* IBL 200) aus verschiedenen Blöcken/Teilclustern (Block 10, Block 15 und Teilcluster 20-A) geteilt wird.

Die vorhergesagte Proteinfamilie der bisher bekannten Immunitätsinhibitoren (InhA1, InhA2 und InhA3) umfasst noch zwei weitere (putative) Metalloproteasen (Abbildung 20). Diese wurden mit InhA4 und InhA5 bezeichnet. Es ist wahrscheinlich, dass die fünf Proteasen in paraloger Beziehung zueinander stehen und durch Genduplikation auseinander hervorgegangen sind. Der Typstamm *B. "manliponensis"* JCM 15802<sup>T</sup> (Block 1) besitzt nur eine Metalloprotease, die keiner der fünf Paralogen eindeutig zugeordnet werden konnte. Hierbei handelt es sich möglicherweise um ein zur Vorläufer-Protease ähnliches Enzym. Die durchschnittlichen Homologien (inklusive einer Standardabweichung), die sich unter den orthologen und zwischen den paralogen Proteinsequenzen ergeben, sind in Tabelle 22 zusammengefasst. Die Ähnlichkeit orthologer Proteine liegt im Durchschnitt zwischen ~ 92 % (InhA1) und ~ 99 % (InhA5), wohingegen die durchschnittliche Homologie paraloger Sequenzen bei ~ 78 % (InhA1 vs. InhA4) im Maximum liegt.



**Abbildung 20 Phylogenie der Gene aus der homologen Familie der Immuninhibitoren**

Ungewurzelte Phylogenie der Gensequenzen aus der vorhergesagten Proteinfamilie der Immuninhibitoren des *B. cereus s.l.* Datensatzes. Berechnet wurde die Phylogenie mit der ML-Methode unter Annahme des GTR+G+I Modells. Die Phylogenie wurde ausschließlich auf Basis vollständiger Gensequenzen rekonstruiert. Blaue Kreise stehen für Äste mit hoher Bootstrap-Unterstützung ( $\geq 70\%$  von 500 Replikaten). Neben den Genen *inhA1* (magenta), *inhA2* (braun) und *inhA3* (dunkles türkis) der drei bereits bekannten Metalloproteasen, sind noch zwei weitere Paraloge Mitglieder der Familie. Diese tragen die Bezeichnungen *inhA4* (orange) und *inhA5* (grün). Zudem verfügt der Typstamm *B. "manliponensis" JCM 15802<sup>T</sup>* über eine putative Metalloprotease, die einen eigenständigen Status einnimmt.

**Tabelle 22 Sequenzähnlichkeiten innerhalb der Proteinfamilie der Immuninhibitoren**

	InhA1	InhA2	InhA3	InhA4	InhA5
<b>InhA1</b>	91,9 ± 4,9				
<b>InhA2</b>	68,2 ± 0,7	95,0 ± 3,5			
<b>InhA3</b>	74,6 ± 1,1	70,5 ± 0,5	95,5 ± 2,8		
<b>InhA4</b>	78,4 ± 1,2	69,0 ± 0,5	75,3 ± 0,8	94,0 ± 3,5	
<b>InhA5</b>	65,6 ± 0,7	59,5 ± 0,4	63,5 ± 0,4	66,4 ± 0,5	98,8 ± 1,2
<b><i>B. "manliponensis" JCM 18502<sup>T</sup></i></b>	71,1 ± 0,6	64,4 ± 0,5	71,0 ± 0,5	72,0 ± 0,4	60,6 ± 0,1

Tabelle umfasst Durchschnittswerte inkl. einer Standardabweichung.

Ähnlichkeitsmatrizen der Clustal Omega Software dienten als Grundlage der Berechnung (siehe Methoden-Abschnitt 14.3. ).

Entgegen dem weitläufigen Vorkommen von *inhA1* (~ 98,5 %), *inhA2* (~ 94 %) und *inhA3* (~ 85 %), ist der Anteil von *inhA4* und *inhA5* tragenden Stämmen im *B. cereus s.l.* Datensatz mit ~ 26 % bzw. ~ 4 % weit geringer. Ins Auge fällt hier, dass alle *B.*

*toyonensis* Stämme neben *inhA1*, *inhA2* und *inhA3* auch *inhA4* besitzen. Allerdings kann weder die Präsenz von *inhA4* allein noch in Verbindung mit den anderen drei Genen als Spezies-spezifisches Charakteristikum gewertet werden. So kommen alle vier CDS zum Beispiel auch in mehreren *B. cereus* s.s (Teilcluster 20-B) und *B. thuringiensis* (Teilcluster 20-A) Isolaten vor. Da *inhA5* bei keinem der Blöcke oder Teilcluster in 100 % der Stämme vorliegt, ist das Gen von vornherein nicht als Spezies Marker geeignet. Nichtsdestotrotz ist auffällig, dass acht von insgesamt elf *inhA5*-positiven Stämmen wiederum der Art *B. toyonensis* angehören.

### 3.3.7. Das Cereulid-Synthetase-Gencluster

Das Cereulid-Synthetase-Gencluster besteht aus sieben proteinkodierenden Genen. *CesA* und *CesB* entsprechen dabei den beiden Modulen der nichtribosomalen Peptidsynthetase (NRPS) zur Bildung des Toxins Cereulid [287].

Das Vorkommen des Genclusters im *B. cereus* s.l. Datensatz ist stark begrenzt. Nur zehn Stämme tragen es in ihrem Genom. Darunter sind drei Stämme (*B. cereus* MC67, *B. cereus* MC118 und *B. cereus* BtB2-4) aus Block 12, sowie sieben Stämme (*B. cereus* AND1407, *B. cereus* AH187, *B. cereus* NC7401, *B. cereus* F1-15, *B. cereus* BDRD-ST26, *B. cereus* MSX-A12 und *B. cereus* IS195) aus Teilcluster 16-A. Eine Eignung des Genclusters als Spezies Marker konnte erwartungsgemäß ausgeschlossen werden.

### 3.3.8. Die Virulenzfaktorprofile der valide beschriebenen *B. cereus* s.l. Arten

Die bisherigen Abschnitte handelten die Verteilung der untersuchten Virulenzfaktoren in einer Klassen-basierten Weise (Enterotoxin-Komplexe, Metalloproteasen, etc.) ab. Im Gegensatz dazu soll dieser Abschnitt in Verbindung mit Tabelle 23 als eine Spezies-zentrierte Zusammenfassung dienen. Es ist jedoch zu beachten, dass nicht jede der valide beschriebenen Arten des *B. cereus* s.l. Datensatz auch eine wohl-definierte Spezies im Sinne dieser Arbeit darstellt (siehe Ergebnis-Abschnitt 3.2.2. ). So können sich bei Ausweitung der Arten *B. mycoides* (Teilcluster 12-D), *B. wiedmannii* (Teilcluster 14-C), *B. anthracis* (Teilcluster 16-C), *B. thuringiensis* (Teilcluster 20-A) und *B. cereus* s.s. (Teilcluster 20-B) auf mehrere Teilcluster gegebenenfalls auch Verschiebungen in den beschriebenen Mustern der Virulenzfaktorprofile ergeben.

Der Art *B. cytotoxicus* (Block 4) fehlt die große Mehrheit der untersuchten Virulenzfaktoren gänzlich. Neben den Genen für die Hauptvirulenzfaktoren Nhe und

CytK verfügt die Spezies ausschließlich über *hlyIII*, *inhA1* und die CDS des Bikarbonat-Transporters. Charakteristisch für *B. cytotoxicus* ist das Auftreten besonderer Gen/Protein Varianten (*cytK-1*, CDS des Bikarbonat-Transporters). Mit einer Ausnahme (Stamm *B. cereus* NVH0597-99 aus Teilcluster 16-C) ist auch das An-/Abwesenheitsprofil der weitläufig vorkommenden Immunitätsinhibitoren (*inhA1*-positiv, *inhA2*-negativ, *inhA3*-negativ) spezifisch für *B. cytotoxicus*. Auch bei *B. pseudomycooides* (Block 9) sind viele der untersuchten Virulenzfaktoren nicht vorhanden. In drei Stämmen fehlt sogar der Enterotoxin-Komplex Nhe und bei mindestens der Hälfte der Stämme ist davon auszugehen, dass HBL nicht funktionstüchtig ist. Die beiden Arten *B. mycooides* (Teilcluster 12-D) und *B. wiedmannii* (Teilcluster 14-C) verfügen über einen hohen Anteil an *hbICDA<sub>a</sub>* kodierenden Stämmen. Zudem besitzt annähernd jeder vierte *B. mycooides* Stamm die genetische Information für mindestens ein Bt-Toxin. Alle *B. anthracis* Stämme (Teilcluster 16-C) besitzen, trotz der weiten Verbreitung des Genes im *B. cereus s.l.* Datensatz, kein *inhA3*. Überdies trägt der klonale *B. anthracis* Komplex eine Nonsense-Mutation im Gen *nheC* sowie eine Frameshift-Mutation im Gen *hlyII*. Die Isolate der Spezies *B. toyonensis* weisen die überwiegende Mehrheit der untersuchten Virulenzfaktoren auf. So besitzen ~ 86 % der Stämme das *nheABC* Operon zusammen mit dem Operon des homologen Komplexes *nheABC<sub>a</sub>*. Das gleiche gilt in Zusammenhang mit *hbICDAB* und *hbICDA<sub>a</sub>*. Darüber hinaus ist neben *inhA1*, *inhA2* und *inhA3* auch *inhA4* in *B. toyonensis* strikt anwesend und die Art umfasst die meisten *inhA5* tragenden Stämme. Die Virulenzfaktorprofile der Isolate aus den verbleibenden zwei valide beschriebenen Arten *B. cereus s.s.* (Teilcluster 20-B) und *B. thuringiensis* (Teilcluster 20-A) ähneln sich stark. Signifikante Unterschiede zeigen sich einzig bei der Verteilung von *hlyII*. Dieses liegt in allen *B. thuringiensis* Stämmen vor, wohingegen es in nur ~ 13 % der *B. cereus s.s.* Isolate anzutreffen ist. Darüber hinaus beherbergen beide Arten den Großteil der Bt-Toxin kodierenden Stämme im Datensatz.

Im Falle von *B. anthracis*, *B. cytotoxicus* und *B. toyonensis* konnten bestimmte Tendenzen in der Verteilung der fünf Immunitätsinhibitoren auffindig gemacht werden. Diese erlauben allerdings keinen eindeutigen und unzweifelhaften Rückschluss auf die jeweilige Art, da die entsprechenden Präsenz/Absenz-Muster auch in den Stämmen anderer *B. cereus s.l.* Spezies vorkommen. Keiner der untersuchten Virulenzfaktoren stellt allein auf Basis seiner bloßen An- bzw. Abwesenheit einen verlässlichen Spezies-diskriminierenden Marker dar. Das Heranziehen einzelner

Virulenzfaktoren als Spezies-definierende Merkmale muss somit als überholt und nicht mehr tragbar angesehen werden.

Tabelle 23 Virulenzfaktorprofile (inkl. Bt-Toxine) der valide beschriebenen *B. cereus s.l.* Arten

Spezies	Verteilung von Virulenzfaktor Genen/Genclustern/Operons (S := prozentualer Anteil kodierender Stämme)			Tendenzen/Auffälligkeiten/Besonderheiten
	S ≥ 95 %	S ≤ 5 %	95 % > S > 5 %	
<b><i>B. cytotoxicus</i></b>				
<b>(Block 4)</b>	<i>nheABC*</i> ,	<i>nheABC<sub>a</sub><sup>Δ</sup></i> , <i>hbICDAB<sup>Δ</sup></i> , <i>hbICDA<sub>a</sub><sup>Δ</sup></i> ,		Sehr viele der untersuchten Virulenzfaktoren nicht vorhanden. Virulenzfaktoren strikt anwesend oder strikt abwesend. Eigene Variante des Bikarbonat-Transporters (Abbildung 15). Eigene <i>cytK</i> Variante (Abbildung 18A).
<u>wohl-definierte Art</u>	<i>cytK*</i> , <i>hlyIII*</i> ,	<i>hlyII<sup>Δ</sup></i> , <i>hlyIIII<sup>Δ</sup></i> ,		
mit 8 Stämmen	<i>inhA1*</i> ,	<i>inhA2<sup>Δ</sup></i> , <i>inhA3<sup>Δ</sup></i> , <i>inhA4<sup>Δ</sup></i> , <i>inhA5<sup>Δ</sup></i> ,		
		Anthrax-Toxin Gene <sup>Δ</sup> , <i>cesHPTABCD<sup>Δ</sup></i> , Bt-Toxin Gene <sup>Δ</sup> , putative Bt-Toxin Gene <sup>Δ</sup>		
<b><i>B. pseudomycooides</i></b>				
<b>(Block 9)</b>	<i>hbICDAB*</i> ,	<i>hbICDA<sub>a</sub><sup>Δ</sup></i> ,	<i>nheABC</i> ~ 73 %, <i>nheABC<sub>a</sub></i> ~ 55 %,	Viele der untersuchten Virulenzfaktoren nicht vorhanden. Drei Stämme ohne <i>nheABC</i> Operon. In sechs Stämmen ist HBL vermutlich nicht funktional.
<u>wohl-definierte Art</u>	<i>hlyIII*</i> , <i>hlyIIII*</i> ,	<i>cytK<sup>Δ</sup></i> , <i>hlyII<sup>Δ</sup></i> ,		
mit 11 Stämmen	<i>inhA1*</i> , <i>inhA2*</i> , <i>inhA3*</i> ,	<i>inhA4<sup>Δ</sup></i> , <i>inhA5<sup>Δ</sup></i> ,		
		Anthrax-Toxin Gene <sup>Δ</sup> , <i>cesHPTABCD<sup>Δ</sup></i> , Bt-Toxin Gene <sup>Δ</sup> , putative Bt-Toxin Gene <sup>Δ</sup>		
<b><i>B. mycooides (B. weihenstephanensis)</i></b>				
<b>(Teilcluster 12-D)</b>	<i>nheABC*</i> ,		<i>nheABC<sub>a</sub></i> ~ 39 %, <i>hbICDAB</i> ~ 61 %, <i>hbICDA<sub>a</sub></i> ~ 83 %,	Hoher Anteil an Stämmen mit <i>hbICDA<sub>a</sub></i> . Ein emetischer Stamm. Dritthöchster relativer Anteil an Bt-Toxin kodierenden Stämmen.
Art mit 23 Stämmen	<i>hlyIII*</i> , <i>hlyIIII*</i> ,	<i>cytK<sup>Δ</sup></i> , <i>hlyII<sup>Δ</sup></i> ,		
	<i>inhA1*</i> , <i>inhA2*</i> , <i>inhA3*</i> ,	<i>inhA4<sup>Δ</sup></i> ,	<i>inhA5</i> ~ 9 %,	
		Anthrax-Toxin Gene <sup>Δ</sup> , <i>cesHPTABCD</i> ,	Bt-Toxin Gene ~ 26 %, putative Bt-Toxin Gene ~ 22 %	

Spezies	Verteilung von Virulenzfaktor Genen/Genclustern/Operons (S := prozentualer Anteil kodierender Stämme)			Tendenzen/Auffälligkeiten/Besonderheiten
	S ≥ 95 %	S ≤ 5 %	95 % > S > 5 %	
<b><i>B. wiedmannii</i></b> (Teilcluster 14-C) Art mit 10 Stämmen	<i>nheABC</i> <sup>*</sup> , <i>hlyIIIIH</i> <sup>*</sup> , <i>inhA1</i> <sup>*</sup> , <i>inhA2</i> <sup>*</sup> , <i>inhA3</i> <sup>*</sup> ,	<i>inhA4</i> <sup>Δ</sup> , <i>inhA5</i> <sup>Δ</sup> , Anthrax-Toxin Gene <sup>Δ</sup> , <i>cesHPTABCD</i> <sup>Δ</sup> , Bt-Toxin Gene <sup>Δ</sup> ,	<i>nheABC</i> <sub>a</sub> ~ 30 %, <i>hbICDAB</i> ~ 90 %, <i>hbICDA</i> <sub>a</sub> ~ 70 %, <i>cytK</i> ~ 20 %, <i>hlyII</i> ~ 70 %, <i>hlyIII</i> ~ 90 %, putative Bt-Toxin Gene ~ 10 %	Hoher Anteil an Stämmen mit <i>hbICDA</i> <sub>a</sub> .
<b><i>B. anthracis</i></b> (Teilcluster 16-C) Art mit 31 Stämmen	<i>nheABC</i> <sup>*</sup> , <i>hlyII</i> , <i>hlyIIIIH</i> <sup>*</sup> , <i>inhA1</i> <sup>*</sup> , <i>inhA2</i> ,	<i>nheABC</i> <sub>a</sub> <sup>Δ</sup> , <i>hbICDA</i> <sub>a</sub> , <i>inhA3</i> <sup>Δ</sup> , <i>inhA4</i> <sup>Δ</sup> , <i>inhA5</i> <sup>Δ</sup> , <i>cesHPTABCD</i> <sup>Δ</sup> ,	<i>hbICDAB</i> ~ 32 %, <i>cytK</i> ~ 48 %, <i>hlyIII</i> ~ 94 %, Anthrax-Toxin Gene ~ 32 %, Bt-Toxin Gene ~ 6 %, putative Bt-Toxin Gene ~ 13 %	Art beinhaltet den klonalen <i>B. anthracis</i> Komplex. Klonaler <i>B. anthracis</i> Komplex trägt Nonsense-Mutation in <i>nheC</i> . Klonaler <i>B. anthracis</i> Komplex trägt Frameshift-Mutation in <i>hlyII</i> . Kein <i>inhA3</i> .
<b><i>B. toyonensis</i></b> (Block 19) wohl-definierte Art mit 22 Stämmen	<i>nheABC</i> <sup>*</sup> , <i>hbICDAB</i> <sup>*</sup> , <i>hlyIII</i> , <i>hlyIIIIH</i> <sup>*</sup> , <i>inhA1</i> <sup>*</sup> , <i>inhA2</i> <sup>*</sup> , <i>inhA3</i> <sup>*</sup> , <i>inhA4</i> <sup>*</sup> ,	<i>hlyII</i> <sup>Δ</sup> , Anthrax-Toxin Gene <sup>Δ</sup> , <i>cesHPTABCD</i> <sup>Δ</sup> , Bt-Toxin Gene, putative Bt-Toxin Gene	<i>nheABC</i> <sub>a</sub> ~ 86 %, <i>hbICDA</i> <sub>a</sub> ~ 86 %, <i>cytK</i> ~ 9 %, <i>inhA5</i> ~ 36 %,	Verfügt über die meisten der untersuchten Virulenzfaktoren. Hoher Anteil an Stämmen mit <i>nheABC</i> und <i>nheABC</i> <sub>a</sub> . Hoher Anteil an Stämmen mit <i>hbICDAB</i> und <i>hbICDA</i> <sub>a</sub> . Alle Stämme besitzen <i>inhA4</i> . Art mit der höchsten Anzahl an <i>inhA5</i> kodierenden Stämmen.

Spezies	Verteilung von Virulenzfaktor Genen/Genclustern/Operons (S := prozentualer Anteil kodierender Stämme)			Tendenzen/Auffälligkeiten/Besonderheiten
	S ≥ 95 %	S ≤ 5 %	95 % > S > 5 %	
<b><i>B. thuringiensis</i></b> (Teilcluster 20-A) Art mit 39 Stämmen	<i>nheABC</i> <sup>*</sup> , <i>hlyII</i> <sup>*</sup> , <i>hlyIII</i> , <i>hlyIIIH</i> <sup>*</sup> , <i>inhA1</i> , <i>inhA2</i> <sup>*</sup> , <i>inhA3</i> ,	<i>nheABC</i> <sub>a</sub> <sup>Δ</sup> ,  <i>inhA5</i> <sup>Δ</sup> , Anthrax-Toxin Gene <sup>Δ</sup> , <i>cesHPTABCD</i> <sup>Δ</sup> ,	<i>hbICDAB</i> ~ 79 %, <i>hbICDA</i> <sub>a</sub> ~ 36 %, <i>cytK</i> ~ 85 %, <i>inhA4</i> ~ 26 %, <p style="text-align: center;">Bt-Toxin Gene ~ 54 %,  putative Bt-Toxin Gene ~ 56 %</p>	Virulenzfaktorprofil ähnelt dem von <i>B. cereus</i> s.s. Im Gegensatz zu <i>B. cereus</i> s.s. strikte Anwesenheit von <i>hlyII</i> . Höchster relativer Anteil an Bt-Toxin kodierenden Stämmen.
<b><i>B. cereus</i> s.s.</b> (Teilcluster 20-B) Art mit 67 Stämmen	<i>nheABC</i> , <i>hbICDAB</i> , <i>cytK</i> , <i>hlyIII</i> , <i>hlyIIIH</i> <sup>*</sup> , <i>inhA1</i> <sup>*</sup> , <i>inhA2</i> <sup>*</sup> , <i>inhA3</i> <sup>*</sup> ,	<i>nheABC</i> <sub>a</sub> ,  <i>inhA5</i> , Anthrax-Toxin Gene <sup>Δ</sup> , <i>cesHPTABCD</i> <sup>Δ</sup> ,	<i>hbICDA</i> <sub>a</sub> ~ 45 %, <i>hlyII</i> ~ 13 %, <i>inhA4</i> ~ 33 %, <p style="text-align: center;">Bt-Toxin Gene ~ 37 %,  putative Bt-Toxin Gene ~ 24 %</p>	Virulenzfaktorprofil ähnelt dem von <i>B. thuringiensis</i> . Im Vergleich zu <i>B. thuringiensis</i> geringes Vorkommen von <i>hlyII</i> . Zweithöchster relativer Anteil an Bt-Toxin kodierenden Stämmen.

Grundlage der Tabelle sind die Virulenzfaktorprofile der Stämme aus dem *B. cereus* s.l. Datensatz (Abbildung 14, Abbildung 16 und Abbildung 17).

<sup>\*</sup>, Virulenzfaktor in allen Stämmen der Spezies vorhanden (S = 100 %).

<sup>Δ</sup>, Virulenzfaktor in keinem Stamm der Spezies vorhanden (S = 0 %).

## VI. DISKUSSION

### 1. Optimierung der PCR-freien Probenvorbereitung und Analyse von Faktoren, die die *de novo* Genomassemblierung beeinflussen

#### 1.1. Faktoren, die die Assemblierung beeinflussen

Eine gute Auswahl an Datenverarbeitungstools sowie die Verwendung qualitativ hochwertiger Read-Daten sind für die Generierung bestmöglicher Genomsequenzen von grundlegender Bedeutung. Sofern der Einfluss unterschiedlicher Parametereinstellungen auf die Qualität der Sequenzierung und der Assemblierung bekannt ist, kann dieses Wissen bei der Weiterentwicklung von Sequenzierungsprojekten helfen.

##### 1.1.1. Der Assembler

Moderne Tools für die Assemblierung von Sequenzen machen Gebrauch von zwei unterschiedlichen algorithmischen Strategien: Overlap-Layout-Consensus (OLC) oder de Bruijn Graphen (dBG). SPAdes ist ein multi-kmer dBG-Assembler [356], wohingegen ABySS und Velvet zu den klassischen dBG-Assemblern zählen [361, 363]. Edena stellt einen OLC-Assembler dar [362]. Da SPAdes im GAGE-B Wettbewerb mit am besten abschnitt [92], ist es nicht verwunderlich, dass dieser Assembler auch den klein angelegten Vergleich dieser Arbeit gewann (Tabelle 8). Mit eingebauten Funktionalitäten für die Read Error und Contig Mismatch Korrektur verkörpert SPAdes den fortschrittlichsten der getesteten Assembler.

##### 1.1.2. Die Insert-Länge

Im Allgemeinen ist der Einfluss der Insert-Länge einer Sequenzierbibliothek auf die Leistungsfähigkeit der Assemblierung abhängig vom verwendeten Assembler. In dieser Arbeit konnten die besten Assemblierungen auf Basis verlängerter Inserts (IS2- und IS3-Bibliotheken) mit den Assemblern SPAdes und ABySS erreicht werden (Tabelle 9 und Tabelle 39). Eine Kombination von Illumina Bibliotheken mit kurzen und (sehr) langen Inserts führte bereits in der Vergangenheit zum Erfolg. Mit Hilfe einer 100-fachen paired-end und 5-fachen mate-pair Bibliothek konnte das *Pseudomonas syringae* pv. *phaseolica* 1448A Genom mit erhöhter Qualität assembliert werden [91]. Der Assembler ALLPATHS-LG, der im GAGE Wettbewerb eine durchgehend starke Leistung zeigte [94], wurde daraufhin entwickelt

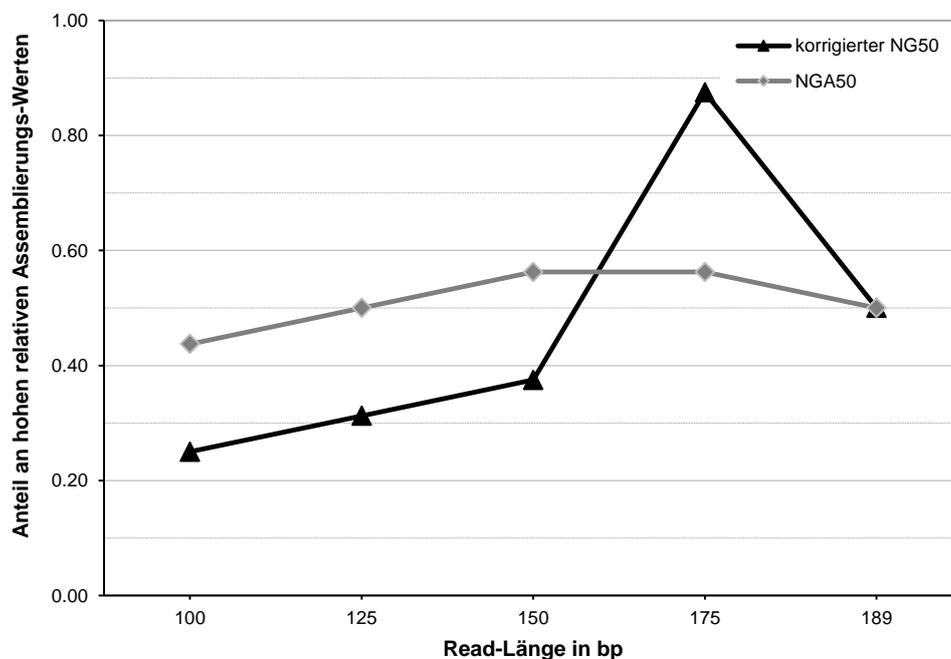
Nutzen aus der Kombination von Illumina paired-end und mate-pair Bibliotheken zu ziehen [415]. Neben der Notwendigkeit zweier Bibliotheken, liegt das Problem derartiger Ansätze bei den mate-pair Bibliotheken, deren Herstellung schwer zu bewerkstelligen ist [416]. Die Anwendung einer paired-end Bibliothek mit erhöhter durchschnittlicher Insert-Länge kann als Alternative dazu angesehen werden. Allerdings bleibt zu beachten, dass der verwendete Assembler dafür ausschlaggebend ist, ob und in welchem Umfang die modifizierte Probenvorbereitung tatsächlich zu überlegenen Assemblierungen führt.

Die Liste zur Verfügung stehender Assembler (alle mit ihren eigenen Vor- und Nachteilen) ist lang. Unglücklicherweise war es im Rahmen dieser Arbeit nicht möglich jeden Assembler zu validieren. Jedoch repräsentieren die hergestellten Bibliotheken einen exzellenten Testdatensatz, der dazu verwendet werden kann herauszufinden, ob in Hinblick auf eine Prolongation der durchschnittlichen Insert-Länge eine Leistungssteigerung bei bestimmten Assemblern erreicht werden kann. Außerdem kann der Testdatensatz von Bioinformatikern dazu genutzt werden die Repeat-Auflösungsfähigkeit bereits bestehender Assembler weiter zu verbessern. Zu diesem Zwecke wurden sämtliche Rohdaten beim NCBI SRA hinterlegt (siehe Methoden-Abschnitt 6.2. ).

### 1.1.3. Die Read-Länge

Im Unterschied zur durchschnittlichen Insert-Länge unterliegt der Einfluss der Read-Länge auf die Assemblierung wesentlich allgemeineren Regeln. In Abbildung 21 ist der Anteil der relativen Assemblierungs-Werte, die in den Bereich  $[0,95, 1]$  fallen, für jede Read-Länge unter Einbeziehung sämtlicher Genome und Assembler gezeigt. Das sich ergebende Gesamtmuster ist im Falle beider Assemblierungs-Metriken ähnlich. Nur wenige der hohen Assemblierungs-Werte werden angesichts der kürzesten Read-Länge (100 bp) erzielt. Dem folgt ein stetiger Anstieg des Anteils hoher Assemblierungs-Werte bis eine Read-Länge von 175 bp erreicht wird. Bei der maximalen Read-Länge von 189 bp fällt der Anteil allerdings wieder ab. Der beschriebene Kurvenverlauf stimmt mit der Erwartung überein. Repetitive Sequenzen verursachen Lücken in Assemblierungen sofern die Read-Länge (oder die Insert-Länge eines Read-Paares) zur Überbrückung eines Repeats nicht ausreicht. Sobald die Read-Länge ansteigt, steigt auch der Anteil eindeutiger Teilsequenzen in der zu assemblierenden Sequenz. Zeitgleich schrumpft die Menge an Teilsequenzen, die zuvor (im Falle einer kürzeren Read-Länge) noch repetitiv

waren [90]. Demnach ist eine Erhöhung der Read-Länge empfehlenswert. Zum jetzigen Zeitpunkt liegt die maximale Read-Länge der Illumina MiSeq Sequenzierplattform bei 2x300 bp mit Insert-Längen von bis zu ~ 1.500 bp. In den meisten bakteriellen Genomen handelt es sich beim 16S rDNA Operon mit 5-7 kpb um das längste repetitive Element [417]. Pacific Biosciences' (PacBio) Technologie der Single Molecule Real Time (SMRT) Sequenzierung ist zur Bildung (sehr) langer Reads und damit zur Überbrückung des 16S rDNA Operons befähigt. Mit den Ansätzen zur Eigen- oder auch hybriden Read Error Korrektur war es möglich die Güte der langen PacBio Reads soweit aufzuwerten, dass Assemblierungen mit einer bis dahin unerreichten Kontiguität entstanden [417, 418]. Jedoch ist die Vorbereitung und Sequenzierung von PacBio Bibliotheken teurer als die von Illumina Bibliotheken [417].



**Abbildung 21 Einfluss der Read-Länge auf die Leistungsfähigkeit von Assemblierungen**

Gezeigt ist der relative Anteil an Assemblierungen mit hohen relativen Assemblierungs-Werten in Abhängigkeit der Read-Länge. Relative Assemblierungs-Werte gelten als hoch, sofern sie einen Wert  $\geq 0,95$  einnehmen. Grundlage für die Berechnung der relativen Anteile sind die Daten aus Tabelle 10, Tabelle 40, Tabelle 43 sowie Tabelle 46.

Diese Abbildung wurde ins Deutsche übertragen und entspricht Abbildung 6 aus der Studie von Huptas *et al.* [342].

Sie unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Doch warum ist ein Abfall des Anteils hoher relativer Assemblierungs-Werte bei einer Read-Länge von 189 bp (Abbildung 21) zu verzeichnen? Wahrscheinlich kann dieses Verhalten mit der Anhäufung von Base Calling Fehlern in Richtung der 3'-Enden der

sequenzierten Reads erklärt werden. Eine solche Akkumulation wirkt störend auf die Assemblierung ein und steht der erhöhten Repeat-Auflösungsfähigkeit dieser Read-Länge entgegen. Es wurde bereits mehrfach nachgewiesen, dass eine der Assemblierung vorausgehende Read Error Korrektur die Leistungsfähigkeit dieser (der Assemblierung) stark positiv beeinflussen kann [419, 420]. Diese Hypothese wird durch die Beobachtung gestützt, dass viele der maximalen relativen Assemblierungs-Werte des Assemblers SPAdes bei einer Read-Länge von 189 bp erreicht wurden (Tabelle 10). Immerhin macht SPAdes Gebrauch von einer Prozedur zur Read Error Korrektur, womit die verbesserte Leistungsfähigkeit des Assemblers bei dieser Read-Länge ein Ergebnis verringerter Base Calling Fehlerraten sein könnte.

#### 1.1.4. Die Sequenziertiefe

Bereits mehrere Studien befassten sich mit der Erkundung der optimalen Sequenziertiefe für die *de novo* Sequenzierung bakterieller Genome [91, 96, 97]. Assemblierungen simulierter Illumina Bibliotheken mit ansteigenden Sequenziertiefen offenbarten N50 Sättigungsgrenzen bei Tiefen zwischen 20- und 40-fach [96]. Dabei konnte bei einer über die Sättigungsgrenze hinausgehenden Anhebung der Sequenziertiefe keine signifikante Zunahme des N50 Wertes mehr erzielt werden. Mit den gängigen Illumina Sequenzierplattformen ist die Sequenzierung prokaryotischer Genome mit hohen Sequenziertiefen (> 100-fach) leicht zu verwirklichen. Vor dem Hintergrund erhöhter finanzieller Kosten bei der Sequenzierung und eines gesteigerten Rechenaufwandes (in Form der benötigten Rechenlaufzeit und des Arbeitsspeicherbedarfes) bei der Assemblierung erscheint die Effizienz derartiger Ansätze jedoch fragwürdig. Untersuchungen dieser Arbeit hinsichtlich der Sequenziertiefe ergaben, dass die Sättigungsgrenzen im Falle des korrigierten NG50 sowie des NGA50 bei SPAdes assemblierten Bibliotheken (IS2 und IS3 mit einer Read-Länge von 189 bp) bei Sequenziertiefen im Bereich von 30- bis 80-fach erreicht wurden (Daten nicht gezeigt). Dies stimmt in etwa mit den Beobachtungen aus der Literatur überein. So saturierten Assemblierungen des *P. syringae* pv. *phaseolica* 1448A Genomes bei einer 100-fachen Sequenziertiefe [91]. Dahingegen konnte bei den meisten Assemblern die optimale Sequenziertiefe zur Assemblierung des *Escherichia coli* MG1655 Genomes auf ~ 50-fach festgelegt werden [97]. Basierend auf den im Rahmen dieser Arbeit gemachten Erfahrungen bildet eine Sequenziertiefe zwischen 30- und 80-fach einen guten Kompromiss

zwischen den anfallenden Kosten einer Sequenzierung und der Etablierung eines optimalen Assemblies.

### 1.2. Optimierte PCR-freie Probenvorbereitungen

Die Prolongation der durchschnittlichen Insert-Länge kann einen positiven Effekt auf die Assemblierung bakterieller Genome haben. Nichtsdestotrotz muss berücksichtigt werden, dass sich eine derartige Verlängerung nachteilig auf die Qualität der Rohdaten auswirkt, was zu einem erheblichen Verlust von Read-Daten während der Qualitätskontrolle führen kann. Da sich dieser nachteilige Effekt mit steigendem GC-Gehalt graduell verstärkt, sollten die Parameter der PCR-freien Probenvorbereitung in Übereinstimmung zum GC-Gehalt des zu sequenzierenden Genomes ausgewählt werden. Demzufolge ist die Herstellung von Bibliotheken mit prolongierten Inserts (IS2 und IS3) nur für AT-reiche Genome zu empfehlen. Unter der Bedingung einer (sehr) hohen Sequenziertiefe kann der Verlust an Read-Daten jedoch ausgeglichen werden. In diesem Fall ist es möglich eine ausreichende Menge an hoch-qualitativen Reads für GC-reiche Genome zu erhalten. Dies geschieht allerdings zu einem höheren Preis, da entsprechend weniger Genome in einem Sequenzierlauf gleichzeitig sequenziert werden können. Sofern IS2- oder IS3-Bibliotheken bei GC-reichen Genomen vonnöten sind, kann auch die Read-Länge verkürzt werden. Da die Qualität der Sequenzierung von Illumina Reads kontinuierlich vom 5'- zum 3'-Ende hin abfällt (Daten nicht gezeigt), wird es einer größeren Anzahl von Read-Paaren gelingen die Filter-Prozedur trotz verlängerter Inserts zu überstehen.

## 2. Taxonomische Richtlinien

*In silico* DDH Vorhersagen gewinnen bei Spezies Neubeschreibungen zunehmend an Bedeutung und werden der experimentellen DDH immer häufiger vorgezogen. Tabelle 24 umfasst alle Spezies Neubeschreibungen der Gattung *Pseudomonas* und der *B. cereus s.l.* Gruppe aus den letzten drei Jahren. Während im Jahr 2015 lediglich ~ 8 % der Neubeschreibungen Gebrauch von phylogenetischen Indices zur Speziesabgrenzung machten, waren es im Jahr 2016 schon ~ 43 % und in 2017 sogar ~ 78 %. Zu den verwendeten Indices gehörten neben einigen ANI-Varianten, wie ANIb, ANIm, OrthoANI, OrthoANLu und gANI [47, 48, 52-54], auch die GBDP [59].

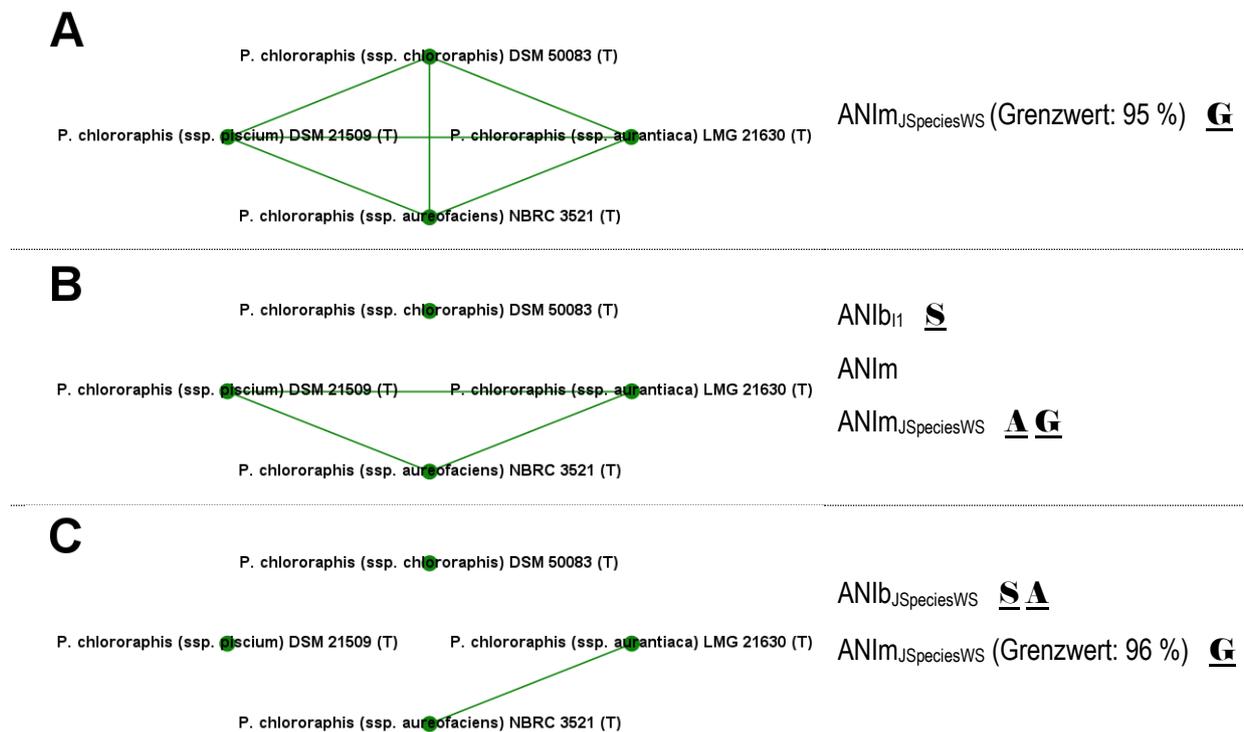
Tabelle 24 *Pseudomonas* und *B. cereus* s.l. Spezies Neubeschreibungen von 2015 bis 2017

Neu beschriebene Spezies <sup>A</sup>	Referenz	DDH Bestimmung		Verwendete Indices zur Speziesabgrenzung (verwendete Software) <sup>URL</sup>
		exp.	in silico	
<i>B. "albus"</i> , <i>B. "luti"</i> , <i>B. "mobilis"</i> und sechs weitere Arten	Liu <i>et al.</i> 2017		x	GBDP (GGDC Web Service) OrthoANIu (ANI Calculator)
<i>P. "aestus"</i>	Vasconcellos <i>et al.</i> 2017	x	x	ANI (JSpecies)
<i>P. "canadensis"</i>	Tambong <i>et al.</i> 2017		x	GBDP (GGDC Web Service) ANIm (JSpecies)
<i>P. "caspiana"</i>	Busquets <i>et al.</i> 2017		x	GBDP (GGDC Web Service) ANib, ANIm (JSpecies)
<i>P. "lactis"</i> , <i>P. "paralactis"</i>	von Neubeck <i>et al.</i> 2017		x	ANib (ANI.pl)
<i>P. "sesami"</i>	Madhaiyan <i>et al.</i> 2017	x		---
<i>P. "tarimensis"</i>	Anwar <i>et al.</i> 2017	x		---
<i>P. "versuta"</i>	See-Too <i>et al.</i> 2017		x	OrthoANI (OAT)
<i>P. "wadenswilerensis"</i>	Frasson <i>et al.</i> 2017		x	GBDP (GGDC Web Service) ANib, ANIm (JSpecies Web Service)
<i>B. wiedmannii</i>	Miller <i>et al.</i> 2016		x	GBDP, ANib (calculate_ani.py)
<i>P. "cerasi"</i>	Kalužna <i>et al.</i> 2016	x	x	GBDP (GGDC Web Service) ANib (ANI calculator)
<i>P. helleri</i> , <i>P. weihenstephanensis</i>	von Neubeck <i>et al.</i> 2016		x	ANib (ANI.pl)
<i>P. "kribbensis"</i>	Chang <i>et al.</i> 2016	x		---
<i>P. oceani</i>	Wang <i>et al.</i> 2016	x		---
<i>P. populi</i>	Anwar <i>et al.</i> 2016	x		---
<i>P. turukhanskensis</i>	Korshunova <i>et al.</i> 2016	x		---
<i>P. alkylphenolica</i>	Mulet <i>et al.</i> 2015		x	GBDP (GGDC Web Service) gANI (Pairwise ANI)
<i>P. asuensis</i>	Reddy & Garcia-Pichel 2015	x		---
<i>P. coleopterorum</i>	Menéndez <i>et al.</i> 2015	x		---
<i>P. donghuensis</i>	Gao <i>et al.</i> 2015	x		---
<i>P. endophytica</i>	Ramírez-Bahena <i>et al.</i> 2015	x		---
<i>P. glareae</i>	Romanenko <i>et al.</i> 2015	x		---
<i>P. granadensis</i>	Pascual <i>et al.</i> 2015	x		---
<i>P. matsuisoli</i>	Lin <i>et al.</i> 2015	x		---
<i>P. salina</i>	Zhong <i>et al.</i> 2015	x		---
<i>P. "songnenensis"</i>	Zhang <i>et al.</i> 2015	x		---
<i>P. yamanorum</i>	Arnau <i>et al.</i> 2015	x		---
<i>P. zhaodongensis</i>	Zhang <i>et al.</i> 2015	x		---

<sup>A</sup> Kennzeichnung nicht valide beschriebener Arten auf Grundlage der LPSN (*Pseudomonas* – Stand April 2017, *B. cereus* s.l. – Stand November 2017)

<sup>URL</sup> Verfügbarkeit und Referenzen der verwendeten Software:  
 ANI Calculator (<https://www.ezbiocloud.net/tools/ani>) [54],  
 ANI calculator (<http://enve-omics.ce.gatech.edu/ani/>) [421],  
 ANI.pl (<https://github.com/chjp/ANI>),  
 calculate\_ani.py ([https://github.com/widow-quinn/scripts/blob/master/bioinformatics/calculate\\_ani.py](https://github.com/widow-quinn/scripts/blob/master/bioinformatics/calculate_ani.py)),  
 GGDC Web Service (<http://ggdc.dsmz.de>) [59],  
 JSpecies (<http://imedea.uib-csic.es/jspecies/index.html>) [48],  
 JSpecies Web Service (<http://jspecies.ribohost.com/jspeciesws>) [51],  
 OAT (<https://www.ezbiocloud.net/tools/orthoani>) [53],  
 Pairwise ANI (<https://img.jgi.doe.gov>) [52]

Im Falle der beiden genomischen Ähnlichkeitsmaße ANIb und ANIm liegt der von der Literatur vorgeschlagene Grenzwert zur Diskriminierung prokaryotischer Arten bei einer Genomähnlichkeit von 95-96 % [47, 48]. Somit befinden sich die im Rahmen dieser Arbeit abgeleiteten Spezies-Cutoffs der Indices ANIb<sub>I1</sub> (95,25 %) und ANIm (95,26 %) in dem von der Literatur empfohlenen Bereich. Abbildung 22 fasst die Vorhersage Index-spezifischer *P. chlororaphis* Spezies-Cluster unter Anwendung von vier unterschiedlichen phylogenomischen Indices sowie der Annahme eines Spezies-Grenzwertes von 95,25 % zusammen. Dabei wurden ausschließlich die Typstämme der vier *P. chlororaphis* Unterarten betrachtet. Grundlage der vier Indices sind zwei unterschiedliche Implementierungen zur Berechnung der genomischen Ähnlichkeitsmaße ANIb und ANIm. Die Spezies-Cluster der Indices ANIb<sub>I1</sub> und ANIm wurden mit dem Skript `caclulate_any.py` (Methoden-Abschnitt 11.2.1. ) berechnet. Die Cluster der Indices ANIb<sub>JSpeciesWS</sub> und ANIm<sub>JSpeciesWS</sub> dahingegen mittels des JSpecies Web Servers (v3.0.17, <http://jspecies.ribohost.com/jspeciesws>). Zusätzlich zum Spezies-Cutoff von 95,25 % sind für den Index ANIm<sub>JSpeciesWS</sub> auch die vorhergesagten Spezies-Cluster, die sich unter den Spezies-Grenzwerten von 95 % bzw. 96 % ergeben, gezeigt.



**Abbildung 22** Abhängigkeiten der Index-spezifischen Spezies-Cluster Vorhersage (*P. chlororaphis*)

Dargestellt ist die Vorhersage Index-spezifischer Spezies-Cluster auf Basis der phylogenomischen Indices ANIb<sub>11</sub>, ANIb<sub>JSpeciesWS</sub>, ANIm und ANIm<sub>JSpeciesWS</sub> bei einem Spezies-Grenzwert von 95,25 %. Im Falle des ANIm<sub>JSpeciesWS</sub> sind zudem die Vorhersageergebnisse bei einem Grenzwert von 95 % bzw. 96 % gezeigt. Zusammengehörige Vorhersagen, die die Abhängigkeit der *in silico* Speziesabgrenzung von der Software, dem Ähnlichkeitsmaß oder dem Spezies-Cutoff verdeutlichen, sind mit den Symbolen “**S**” (für Software), “**A**” (für Ähnlichkeitsmaß) und “**G**” (für Grenzwert) gekennzeichnet. Je nach Software, Index und Grenzwert ergeben sich auf Basis der vorhergesagten Spezies-Cluster unter Umständen unterschiedliche taxonomische Rückschlüsse: **A** *P. chlororaphis* als eine Nomenspezies mit vier Unterarten (Dies entspricht dem aktuellen taxonomischen Status der Art nach LPSN). **B** Vorschlag zur Erhebung von *P. chlororaphis* ssp. *aurantiaca* in den Rang einer eigenständigen Nomenspezies. **C** Vorschlag der Veränderung des Status von *P. chlororaphis* ssp. *aurantiaca* und *P. chlororaphis* ssp. *piscium* hin zu jeweils eigenständigen Nomenspezies.

Aus Abbildung 22 wird ersichtlich, dass die Vorhersage der Spezieszugehörigkeit prokaryotischer Isolate abhängig von der verwendeten Software (Symbol “**S**”), dem herangezogenen Algorithmus zur Abschätzung der genomischen Ähnlichkeit (Symbol “**A**”) sowie dem angesetzten Spezies-Grenzwert (Symbol “**G**”) ist. Gemäß der aktuellen Taxonomie handelt es sich bei *P. chlororaphis* um eine Nomenspezies bestehend aus vier Unterarten (Abbildung 22A). Der Einfluss der jeweiligen Implementierung (Software) auf die Speziesabgrenzung wird anhand der unterschiedlichen Spezies-Cluster Vorhersagen des ANIb<sub>11</sub> und ANIb<sub>JSpeciesWS</sub> deutlich. Im Falle des ANIb<sub>11</sub> würde es zur Aufteilung von *P. chlororaphis* in zwei potentielle Nomenspezies kommen, wohingegen die ANIb<sub>JSpeciesWS</sub> Vorhersage drei potentielle Nomenspezies prognostiziert (vgl. Abbildung 22B mit Abbildung 22C).

Derartige Diskrepanzen können in einer vom ursprünglichen Algorithmus [47] abweichenden (oder sogar fehlerhaften) Implementierung des Ähnlichkeitsmaßes (hier ANIb) liegen. Zudem können auch Implementierungsdetails (wie z.B. sich unterscheidende Standardparametereinstellungen) zu inkongruenten Vorhersageergebnissen führen. Ein Beispiel voneinander abweichender Spezies-Cluster Vorhersagen, das auf die Verwendung unterschiedlicher Ähnlichkeitsmaße zurückzuführen ist, liefern die Indices ANIb<sub>JSpeciesWS</sub> und ANIm<sub>JSpeciesWS</sub>. Entgegen der Vorhersage dreier potentieller Nomenspezies (ANIb<sub>JSpeciesWS</sub>) würde es beim ANIm<sub>JSpeciesWS</sub> zur Aufteilung von *P. chlororaphis* in zwei potentielle Nomenspezies (vgl. Abbildung 22B mit Abbildung 22C) kommen. Entsprechend ihrer algorithmischen Ausrichtung (hier die Wahl des Alignment-Programmes) ist davon auszugehen, dass unterschiedliche Ähnlichkeitsmaße sich bei der Berechnung des Ähnlichkeitswertes eines Genompaares auf teils unterschiedliche Genombereiche stützen. Auch der Anteil zweier verglichener Genome, der in die Berechnung des Ähnlichkeitswertes mit einfließt, kann unterschiedlich hoch ausfallen. Letztendlich hat beides Auswirkungen auf die Abschätzung der Genomähnlichkeit. Die Abhängigkeit der *in silico* DDH vom jeweils angenommenen Spezies-Grenzwert ist offensichtlich und kann anhand der drei ANIm<sub>JSpeciesWS</sub> Vorhersagen leicht nachvollzogen werden. Je nach Stringenz des zugrunde gelegten Spezies-Grenzwertes kommt es dabei zur Vorhersage von bis zu drei potentiellen Nomenspezies (vgl. Abbildung 22A, Abbildung 22B und Abbildung 22C).

Vor dem Hintergrund einer stetig steigenden Anzahl an zur Verfügung stehenden *in silico* DDH-Varianten (GBDP, ANIb, ANIm, etc.) und Software-Implementierungen für ihre Berechnung [48, 51-55, 59, 421] stellt sich die Frage, ob eine Beschränkung auf bestimmte Software-Pakete und DDH-simulierende Algorithmen stattfinden sollte, um die Beständigkeit neu beschriebener Arten zu wahren? In diesem Zusammenhang sollte außerdem über die Einführung von klar definierten Spezies-Grenzwerten bei den diversen ANI-Implementierungen nachgedacht werden. Im Falle der GBDP wurde ein strikter Grenzwert ( $dDDH \geq 70\%$ ) von vornherein festgelegt [59]. Dies scheint der taxonomischen Nutzung des Index jedoch nicht entgegenzustehen. Von den insgesamt elf Neubeschreibungen aus Tabelle 24, die Gebrauch von einer *in silico* DDH Vorhersage machten, verwendeten immerhin sieben (~ 64 %) die GBDP. Die Problematik, die mit dem Einsatz vielfältiger Software-Implementierungen und Algorithmen zur Abschätzung genomischer Ähnlichkeiten einhergehen kann

(Abbildung 22), scheint noch nicht erkannt worden zu sein. Bisher ruft die gängige Literatur einzig zur Gewinnung und Bereitstellung genomischer Sequenzdaten sowie zur Nutzung von *in silico* DDH-Prozeduren zum Zwecke der taxonomischen Charakterisierung neuer prokaryotischer Arten auf [3, 12, 61, 78]. Reklassifizierungen infolge eines Genomvergleiches von *Pseudomonas* oder *B. cereus* s.l. Typstämmen, wurden jedoch bereits vorgeschlagen, obgleich es sich dabei um Arten, die ursprünglich nicht mit Hilfe von *in silico* DDH-Abschätzungen beschrieben wurden, handelt [413, 422]. Angesichts der Kosteneffizienz von Genomsequenzierungen [37, 43], der Empfehlung führender Taxonomie-Journale zur Sequenzierung des Typstammes im Zuge von Spezies Neubeschreibungen [3] und den Anstrengungen von Initiativen zur Sequenzierung von Typstammgenomen (wie z.B. GEBA [79]) ist zu erwarten, dass eine zügige Verdrängung der arbeitsaufwendigen und fehleranfälligen experimentellen DDH durch die *in silico* DDH bei Neubeschreibungen stattfinden wird. Aus der klein angelegten Analyse der *Pseudomonas* und *B. cereus* s.l. Neubeschreibungen in Tabelle 24 konnte ein derartiger Trend jedenfalls unzweifelhaft abgeleitet werden. Zur Vermeidung taxonomischer Inkonsistenzen und der Notwendigkeit zukünftiger Reklassifizierungen, die allein auf die Verwendung unterschiedlicher Software-Implementierungen, genomischer Ähnlichkeitsmaße und/oder Spezies-Grenzwerte zurückzuführen sind, sollte die Gemeinschaft der prokaryotischen Taxonomen die Ausarbeitung von allgemein verbindlichen Richtlinien in Angriff nehmen.

### 3. Speziesabgrenzung in der Gattung *Pseudomonas*

#### 3.1. Speziesabgrenzung via MLSA

Das MLSA Schema von Mulet *et al.* [25] wurde schon sehr häufig im Rahmen taxonomischer und phylogenetischer Studien angewendet. Mit Hilfe dieses Schemas wurde die phylogenetische Struktur der Gattung *Pseudomonas* erstmals im Detail beschrieben [25] und unter Einbeziehung einer zunehmenden Anzahl an Typstämmen bereits zweimal aktualisiert und erweitert [125, 128]. Überdies konnte die phylogenetische Zugehörigkeit von 33 *P. putida* Biovar A und B Stämmen aufgeklärt [129] und vier neue *P. stutzeri* Genomovare beschrieben werden [130, 136]. Außerdem stützten sich mehr als 50 % der *Pseudomonas* Neubeschreibungen (Tabelle 24), die Gebrauch von mindestens zwei konkatenierten Housekeeping-

Genen zur phylogenetischen Charakterisierung der untersuchten Isolate machten, auf dieses Schema.

Ursprünglich wurde der Grenzwert zur Diskriminierung von *Pseudomonas* Arten bei einer MLSA Ähnlichkeit von 97 % angesetzt [25]. Dieser Grenzwert konnte basierend auf den Genomvergleichen von 112 Stämmen (darunter lediglich zwei Typstämme) etwas später nochmals bekräftigt werden [125]. Im Gegensatz zur letzteren Studie erfolgte die Festsetzung des MLSA Spezies-Grenzwertes in dieser Arbeit auf Grundlage des paarweisen Vergleiches von 156 Typstammgenomen. Somit ist der angepasste MLSA Spezies-Grenzwert von 97,84 % sowohl in qualitativer wie auch quantitativer Hinsicht weitaus besser begründet. Allerdings muss beachtet werden, dass die Berechnung der MLSA Ähnlichkeit nicht identisch mit derjenigen aus den vorhergehenden beiden Studien ist. Entgegen der Verwendung umgerechneter Jukes-Cantor (JC) Distanzen (Ähnlichkeit :=  $100 - \text{JC Distanz}$ ) [25, 125] wurde in dieser Arbeit der prozentuale Anteil übereinstimmender Basen im multiplen Sequenzalignment (= Sequenzähnlichkeit) als Maß für die Ähnlichkeit zweier Nukleinsäuresequenzen herangezogen. Die Höhe der phylogenetischen Distanz eines Sequenzpaares ist immer abhängig vom zugrunde gelegten Nukleotid-Substitutionsmodell. Anstelle des JC Modells wurde allerdings das GTR+G+I Modell als das am besten geeignete Modell für das multiple MLSA Sequenzalignment des *Pseudomonas* Datensatz ermittelt (siehe Methoden-Abschnitt 10.4.1. ). Da jedem genomischen Ähnlichkeitsmaß (ANiB, ANIm, etc.) eine Abschätzung der durchschnittlichen genomweiten Sequenzähnlichkeit zugrunde liegt, ist die Anwendung von MLSA Sequenzähnlichkeiten zur Speziesabgrenzung zudem theoretisch fundierter.

Stellenweise wurden in der Literatur unterschiedliche MLSA Grenzwerte zur Diskriminierung von *Pseudomonas* Spezies angewendet. Bei der Abgrenzung von *Pseudomonas* Nomenspezies kam dabei der Grenzwert von 97 % zum Einsatz [25, 125, 128, 129], wohingegen für die Charakterisierung und Beschreibung von *P. stutzeri* Genomovaren auf einen Cutoff von 95,2 % zurückgegriffen wurde [130, 136]. Das Genomovar Konzept wurde erstmalig zum Zwecke der Aufklärung des taxonomischen Status der Nomenspezies *P. stutzeri* eingeführt [423, 424]. Dabei handelt es sich um Gruppen von Isolaten, die zwar auf genotypischer Basis unterschieden werden können (DDH < 70 %), jedoch phänotypisch ähnlich sind [425]. Demzufolge repräsentieren Genomovare Genomospezies, die einer

Nomenspezies zugeordnet sind. Sobald bei einer Genomospezies differenzierende phänotypische Merkmale festgestellt werden, kann eine Benennung hin zur eigenständigen Nomenspezies vorgenommen werden [9, 426]. Beispielsweise wird das ehemalige *P. stutzeri* Genomovar 6 heute durch die Nomenspezies *P. balearica* repräsentiert [427]. Vor dem Hintergrund der Transformierbarkeit einer Genomospezies in eine Nomenspezies erscheint die Verwendung zweier unterschiedlicher MLSA Grenzwerte als höchst fragwürdig. Denn in dem Moment in dem aus einem *P. stutzeri* Genomovar eine *Pseudomonas* Nomenspezies wird, treten in Bezugnahme auf die genotypische Kohärenz der Art (aufgrund des erhöhten Grenzwertes) restriktivere Maßstäbe in Kraft. Unter dem "neuen" Grenzwert (von 97 %) kann eine Bildung von mehreren zueinander disjunkten genotypischen Gruppen von Isolaten nicht ausgeschlossen werden. In diesem Fall würde sich ein Genomovar in mindestens zwei Nomenspezies aufteilen, sofern für jede Gruppe diskriminative phänotypische Merkmale gefunden werden können. Einerseits führt dies zu einem Widerspruch in sich, denn ursprünglich wurde das Genomovar ja gerade eben auf Grundlage seiner genotypischen Kohärenz definiert. Andererseits hat dies auch Auswirkungen auf die Praxis. Wird bei der Überführung einer Genomospezies in eine Nomenspezies eine derartige Grenzwert-vermittelte genotypische Aufteilung nicht sofort bemerkt, ist davon auszugehen, dass zukünftige Reklassifizierungen zur Wahrung eines in sich stimmigen taxonomischen Systems vonnöten sein werden. Wie bereits in Diskussions-Abschnitt 2. gefordert, untermauern die hiesigen Ausführungen nochmals die Notwendigkeit von eindeutig definierten Spezies-Grenzwerten bei taxonomischen Studien.

16S rDNA Sequenzen und die Housekeeping-Gene *gyrB*, *rpoB* und *rpoD* wurden bereits vor der Entwicklung des drei- bzw. vier-Gen MLSA Schemas [25] für die phylogenetische Analyse der Gattung *Pseudomonas* genutzt [118, 119, 121, 123]. Außerdem zählen die vier Gene zu denjenigen, die für gewöhnlich in MLSA Schemata anzufinden sind [19]. Insofern stellt sich die Frage, inwieweit es sich bei diesem Schema tatsächlich um ein optimales Schema für die Speziesabgrenzung in der Gattung *Pseudomonas* handelt. Bereits bekannt ist, dass das 16S rDNA Gen nur über eine geringe intraspezifische Auflösungskraft verfügt [118, 119, 121]. Darüber hinaus kann das Gen in multipler Kopienzahl (Paraloge) im Genom vorliegen. Die Sequenzierung nicht identischer 16S rDNA Amplifikate kann zur Generierung einer digitalen Hybrid-Sequenz führen und im Falle der Extraktion mehrerer nicht übereinstimmender 16S Sequenzen aus Genomassemblierungen ist fraglich welche

Sequenz letztendlich verwendet werden sollte. Das Gen des MLSA Schemas mit der höchsten intraspezifischen Trennschärfe ist *rpoD* gefolgt von *gyrB* und *rpoB* [25]. In der Arbeit von Mulet *et al.* basierte die Abschätzung des Grades der zwischenartlichen Unterscheidungsfähigkeit allerdings einzig auf dem gegenseitigen Vergleich der vier Gene. Ein Abgleich der phylogenetischen Distanzen, die sich beim paarweisen Vergleich der 107 Typstämme für jedes Gen ergaben, mit phylogenomischen Ähnlichkeiten blieb aufgrund der Nichtverfügbarkeit von genomischen Sequenzdaten aus. In einer vorausgehenden Studie konnten jedoch ausgehend von 49 paarweisen Genomvergleichen aus über 100 weit verbreiteten proteinkodierenden Housekeeping-Genen 32 vielversprechende Kandidaten für die bakterielle Speziesabgrenzung ausfindig gemacht werden [20]. Unglücklicherweise ist keines der vier MLSA Gene unter den Kandidaten anzufinden. Überdies macht das Species Identification Tool (specl) Gebrauch von 40 universellen Markergenen [428, 429], um prokaryotische Organismen in Spezies-Cluster zu gruppieren [430]. Von den vier MLSA Genen gehört einzig *rpoB* zu den universellen Markern.

Obgleich davon auszugehen ist, dass das MLSA Schema von Mulet *et al.* keine optimale Lösung zur Speziesabgrenzung darstellt, ist zu hinterfragen, ob die Entwicklung und Etablierung eines "optimalen" Schemas sowohl aus wissenschaftlicher wie auch wirtschaftlicher Sicht noch rentabel genug ist. Hinsichtlich der Geschwindigkeit und Kosteneffizienz mit der prokaryotische Genome zum gegenwärtigen Zeitpunkt sequenziert und assembliert werden können sowie der Verfügbarkeit von *in silico* DDH Prozeduren zur Speziesabgrenzung ist die Kosten-Nutzen-Relation als eher ungünstig anzusehen. Diesbezüglich stellt die weitergehende Verwendung des Schemas unter Berücksichtigung des verfeinerten Spezies-Grenzwertes einen guten Kompromiss zur Ausarbeitung eines gänzlich neuen Schemas dar. Nichtsdestotrotz ist zur Kenntnis zu nehmen, dass selbst unter dem angehobenen Grenzwert die Unterscheidung bestimmter *Pseudomonas* Arten nicht möglich ist (Tabelle 18).

### 3.2. Reklassifizierung beschriebener Arten

Infolge des phylogenetischen (MLSA) und phylogenomischen (ANi<sub>b1</sub>, ANi<sub>b2</sub>, ANi<sub>m</sub>, gANI) Vergleiches der *Pseudomonas* Typstämme (aus Tabelle 48) resultierten bei 18 Typstammpaaren Index-übergreifende (deckungsgleiche) Speziesvorhersagen, die der aktuellen Taxonomie (LPSN April 2017) entgegenstanden und zum Vorschlag von zwölf Reklassifizierungen führten. Diese sind in Tabelle 17 zusammengefasst.

Die genomische Kohärenz der Arten *P. amygdali*, *P. ficuserectae*, *P. meliae* und *P. savastanoi* wurde mittels experimenteller DDH bereits nachgewiesen [431]. Darüber hinaus kam es auch schon zur Unterbreitung des Vorschlages einer Behandlung von *P. tuomuerensis* als späterem heterotypischem Synonym von *P. flexibilis* [422]. Im Falle der Neubeschreibungen von *P. panacis*, *P. psychrotolerans* und *P. "sihuiensis"* wurden die Typstämme der Arten *P. marginalis*, *P. oryzihabitans* und *P. chengduensis* bei der taxonomischen Analyse nicht mitberücksichtigt [432-434] und die Charakterisierung von *P. fuscovaginae* basierte auf rein phänotypischen Merkmalen [435]. Die ehemaligen Typstämme *P. perfectomarina* ATCC 14405 und *P. chloritidismutans* AW1 sind Mitglieder der *P. stutzeri* Genomovare 2 und 3 [403, 405]. Beide Isolate weisen im Vergleich zum *P. stutzeri* Typstamm einen experimentellen DNA-DNA Reassoziierungsgrad  $< 70\%$  auf [401, 404], womit ihr Status als eigenständige Genomospezies außer Frage steht. Die Neubeschreibung von *P. kunmingensis* beruht unter anderem nur auf dem Vergleich mit dem *P. stutzeri* Typstamm [436]. Da der ehemalige *P. chloritidismutans* Typstamm bei der Charakterisierung der Nomenspezies nicht miteinbezogen wurde, kann eine Zugehörigkeit des *P. kunmingensis* Typstammes zum *P. stutzeri* Genomovar 3 nicht ausgeschlossen werden. Mit Ausnahme der Stämme *P. zeshuii* KACC 15471<sup>T</sup> und *P. luteola* NBRC 103146<sup>T</sup>, die einen experimentellen DDH Wert von  $\sim 29\%$  aufweisen [437], untermauern alle bisherigen Ausführungen den hohen Verlässlichkeitsgrad der vorgeschlagenen Reklassifizierungen aus Tabelle 17. Zu beachten ist jedoch, dass auf einen formalen Antrag zur Reklassifizierung der betroffenen Arten verzichtet wird, da im Rahmen einer polyphasischen Taxonomie, die zum momentanen Zeitpunkt (noch) gängige Praxis ist, wichtige Voraussetzungen, wie z.B. der Beweis phänotypischer Kohärenz (bzw. Inkohärenz im Falle von *P. stutzeri*, *P. chloritidismutans* und *P. perfectomarina*), durch diese Arbeit nicht erbracht wurden.

### 3.3. Anzahl potentieller neuer Arten

Aus Tabelle 19 geht hervor, dass die geschätzte Anzahl potentieller neuer *Pseudomonas* Genomospezies auf ein Minimum von 57 Taxa (MLSA Vorhersage) festgesetzt werden kann. Hinzu kommt, dass bei jedem der fünf Indices (MLSA, ANI<sub>b1</sub>, ANI<sub>m</sub>, ANI<sub>b2</sub> und gANI) die überwiegende Mehrheit der neuen Arten durch nur einen einzigen Stamm (Singletons) vertreten ist. Dies lässt darauf schließen, dass der hiesige *Pseudomonas* Datensatz die in der Natur vorkommende Diversität der Gattung nur unzureichend wiedergibt. Angesichts des ubiquitären und weltweiten

Vorkommens von *Pseudomonas* Isolaten [104, 105] sowie der immensen metabolischen und physiologischen Vielfalt der Gattung [151, 152] ist dies nicht weiter verwunderlich. Rarefaction und Collector's Curve Analysen ergaben, dass Dutzende *P. stutzeri* Genomovare zu erwarten sind [130] und dass in der *P. fluorescens* Gruppe Hunderte von Genomospezies-Clustern vermutet werden können [132]. Infolge eines Spezies Reichtums, der in die Tausende gehen könnte, ist in Zukunft mit einer weiteren Revision der Gattung zu rechnen.

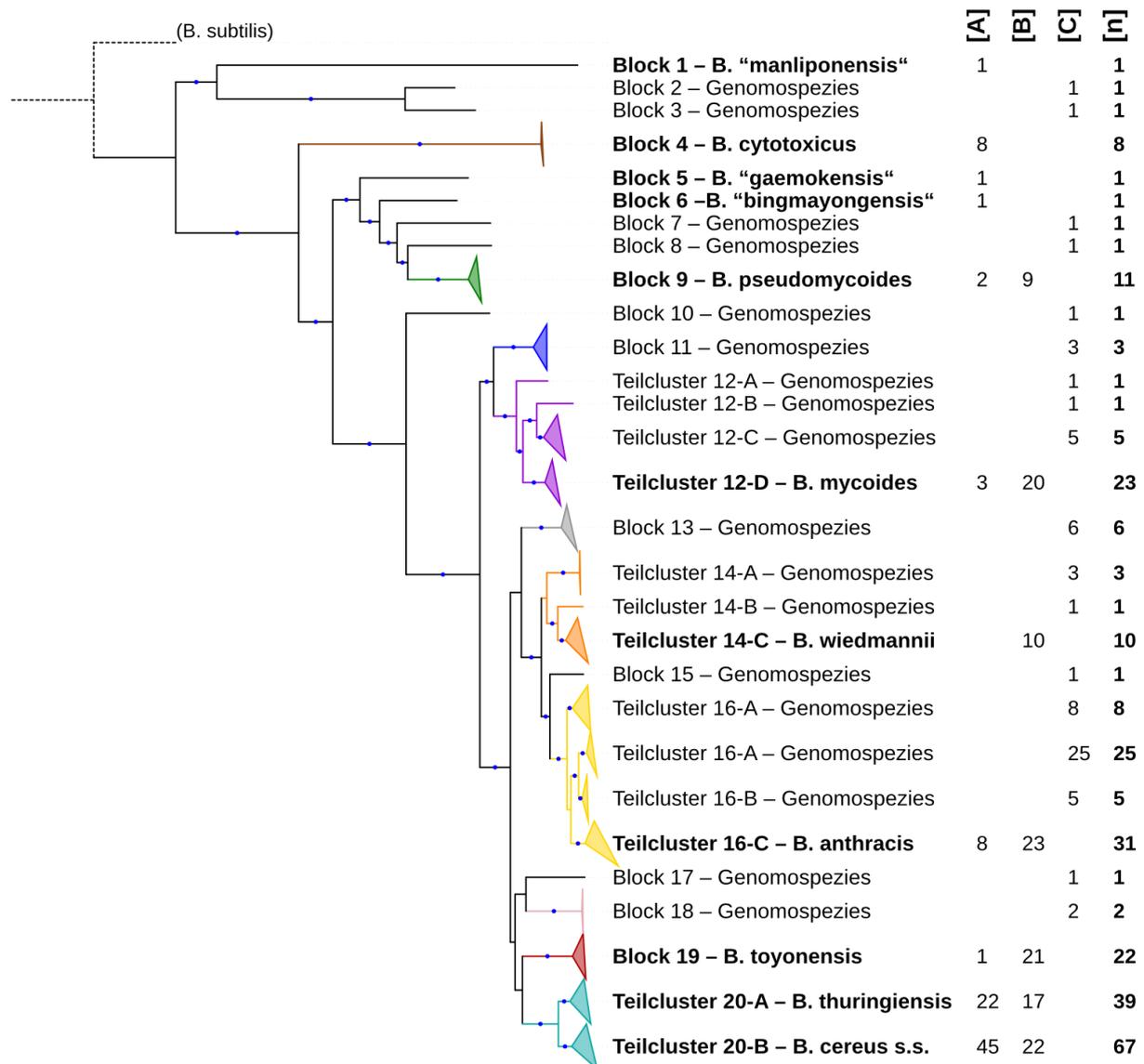
#### 4. Taxonomie der *B. cereus* s.l. Gruppe

##### 4.1. Speziesabgrenzungen

Aufgrund ihrer medizinischen und ökonomischen Bedeutung werden viele *B. cereus* s.l. Isolate noch heute entsprechend ihrer phänotypischen Eigenschaften, wie z.B. dem verursachten Erkrankungsbild (*B. anthracis*, *B. cereus* s.l., emetische *B. cereus* s.l.), der Fähigkeit zur Produktion bestimmter Toxine (*B. anthracis*, *B. thuringiensis*, emetische *B. cereus* s.l.), dem Wachstum bei niedrigen Temperaturen (*B. mycoides*/*B. weihenstephanensis*) oder der Koloniemorphologie (*B. mycoides*, *B. pseudomycoides*), benannt. Mitunter erfolgt die Namensgebung auch in Abhängigkeit vom jeweiligen Untersuchungskontext. Beispielsweise verwenden Studien, die sich mit dem Verderbspotential von psychrotoleranten Sporenbildnern bei Lebensmitteln befassen, häufig das 16S rDNA oder *rpoB* Gen zur Identifizierung von Isolaten, womit eine Charakterisierung bis auf Speziesebene nicht immer möglich ist und oft auch nicht angestrebt wird [202, 205]. So wurde der *B. wiedmannii* Typstamm anfänglich nur als ein Mitglied der *B. cereus* s.l. Gruppe klassifiziert [169]. Vor allem die historisch gewachsene Abhängigkeit der Spezieszugehörigkeit vom Phänotyp hat dazu geführt, dass die Bezeichnung sehr vieler *B. cereus* s.l. Stämme nicht die tatsächliche phylogenetische/phylogenomische Stellung der Isolate widerspiegelt [26, 253, 438]. Fast 83 % von den in dieser Arbeit verwendeten 281 Stämmen werden allein den Arten *B. cereus* s.s. (~ 63 %) und *B. thuringiensis* (~ 20 %) zugeordnet, wobei lediglich 106 Isolate des Datensatzes (~ 38 %) in Wirklichkeit Mitglieder der beiden entsprechenden monophyletischen Gruppen sind. Um derartige Fehlbenennungen zukünftig zu vermeiden, sollte die *B. cereus* s.l. Taxonomie strikt nach Genomospezies-Clustern ausgerichtet werden. Im Rahmen dieser Arbeit wurden derartige Cluster (basierend auf den ANI Speziesvorhersagen) in Form von Blöcken und Teilclustern für den *B. cereus* s.l. Datensatz bereits vorgeschlagen

(Abbildung 12 und Abbildung 13) und sind in Verbindung mit der MLSA Phylogenie in Abbildung 23 zusammengefasst. Die 28 Genomospezies-Cluster teilen sich in elf beschriebene Nomenspezies und 17 potentielle neue *B. cereus s.l.* Arten auf. Dies ist deckungsgleich mit den Beobachtungen einer ähnlich angelegten Studie, bei der mittels des phylogénomischen Index GBDP 19-20 putative neue *B. cereus s.l.* Spezies postuliert wurden [438]. Unter der Annahme, dass alle 28 Genomospezies jeweils eine Nomenspezies repräsentieren, muss in Zukunft mit einer Umbenennung von mehr als zwei Dritteln der Stämme des hiesigen Datensatzes gerechnet werden. Ungeachtet dessen ist zu erwarten, dass noch viele bis dato unbekannte Arten in der *B. cereus s.l.* Gruppe existieren. So kam es nach Abschluss der vorliegenden Analyse zur Neubeschreibung von neun Arten [439].

Tree scale: 0.1



**Abbildung 23 Vorhergesagte Genomospezies-Cluster im *B. cereus* s.l. Datensatz**

Die MLSA Phylogenie des *B. cereus* s.l. Datensatzes wurde unter Annahme des GTR+G+I Substitutionsmodells mit der ML-Methode rekonstruiert. Äste mit hoher Bootstrap-Unterstützung ( $\geq 70\%$  von 500 Replikaten) sind durch blaue Kreise gekennzeichnet. Die Außengruppe *B. subtilis* ATCC 6051<sup>T</sup> wurde zur Wurzelung des Baumes verwendet. Die phylogenetischen Distanzen zur Außengruppe wurden künstlich auf 0,1 Substitutionen pro Alignmentposition verringert. Auf Grundlage der ANI Speziesvorhersagen ergeben sich 28 Genomospezies-Cluster (Blöcke und Teilcluster), die als kondensierte Knoten im Baum vorliegen. Die elf Genomospezies-Cluster der bereits bekannten Arten sind mit dem Namen der jeweiligen Nomenespezies versehen und durch Fettdruck hervorgehoben. Die abschließenden Spalten zeigen **[A]** die Anzahl an richtig benannten Stämmen, **[B]** die Anzahl notwendiger Umbenennungen in eine bekannte Nomenespezies, **[C]** die Anzahl notwendiger Umbenennungen in eine potentielle neue Spezies sowie **[n]** die Gesamtanzahl an Stämmen im jeweiligen Genomospezies-Cluster.

#### 4.2. *B. cereus s.l.* Virulenzfaktoren als Spezies-diskriminierende Marker

Hinsichtlich ihres Spezies-diskriminierenden Potentials konnten signifikante Assoziationen zwischen den drei bekannten Immuninhibitoren (InhA1-3) und den zwei Arten *B. cytotoxicus* sowie *B. anthracis* aufgedeckt werden (Abbildung 14). Zunächst ist festzuhalten, dass trotz der weiten Verbreitung der drei Metalloproteasen Gene die Spezies *B. cytotoxicus* (Block 4) einzig über *inhA1* verfügt. Dahingegen ist die Art *B. anthracis* (Teilcluster 16-C) *inhA3*-defizient. In Kombination mit den jeweiligen beiden anderen Genen ergeben sich jedoch An-/Abwesenheitsprofile, die nur unter gewissen Einschränkungen als Spezies-spezifisch gewertet werden könnten. So kommt das zu *B. cytotoxicus* gehörige Profil (*inhA1*-positiv, *inhA2*-negativ, *inhA3*-negativ) auch im Stamm *B. cereus* NVH0597-99 aus Teilcluster 16-C vor. Das mit *B. anthracis* assoziierte Muster (*inhA1*-positiv, *inhA2*-positiv, *inhA3*-negativ) ist sogar in drei Stämmen (*B. mycoides* BHP, *B. cereus* F528-94 und *B. thuringiensis* IBL 200) ohne Zugehörigkeit zur Art präsent.

Neben den drei bekannten Immuninhibitoren konnte die Existenz zweier weiterer Metalloproteasen (InhA4 und InhA5) durch diese Arbeit nachgewiesen werden (Abbildung 20). Angesichts der geringen Häufigkeit von *inhA4* (~ 26 %) und *inhA5* (~ 4 %) im *B. cereus s.l.* Datensatz (Abbildung 14) sticht das gehäufte Vorkommen der Gene in den Stämmen der Art *B. toyonensis* (Block 19) ins Auge. Dementsprechend trägt jedes der 22 Isolate *inhA4* und von insgesamt elf *inhA5* kodierenden *B. cereus s.l.* Stämmen im Datensatz entfallen allein acht auf diese Spezies. Außerdem ist auffällig, dass der Großteil der *B. toyonensis* Isolate (~ 73 %) neben den Genclustern für die Drei-Komponenten-Toxine Nhe und HBL auch über die Operons der homologen Enterotoxin-Komplexe Nhe<sub>a</sub> und HBL<sub>a</sub> verfügt. Untersuchungen mit dem Stamm *B. cereus* MGBC 145 ergaben, dass HBL und HBL<sub>a</sub> über eine vergleichbare Effektivität bei der Erhöhung der vaskulären Permeabilität und ein ähnliches hämolytisches Potential verfügen [306]. Darüber hinaus wurde in der Studie von Böhm *et al.* die Transkription der Operons *nheABC* und *nheABC<sub>a</sub>* bei zwei von drei getesteten Stämmen festgestellt [26]. Außerdem konnte gezeigt werden, dass die drei Metalloproteasen InhA1, InhA2 und InhA3 während der Infektion im Insektenmodell koexprimiert werden und dass die Virulenz einer *inhA*-defizienten *B. thuringiensis* Dreifachmutante stark eingeschränkt ist [331]. Ob kumulative, kooperative oder sogar synergistische Wechselbeziehungen zwischen den Enterotoxin-Komplexen (Nhe, HBL) und ihren homologen Toxinen (Nhe<sub>a</sub>, HBL<sub>a</sub>)

existieren und inwieweit derartige Beziehungen auch unter den Immulinhibitoren (InhA1-5) zu finden sind, sollte zukünftig intensiver untersucht werden. Jedenfalls ist bekannt, dass die Anwesenheit von SM-PLC die HBL bedingte Lyse von Erythrozyten verstärkt (synergistischer Effekt) [440]. Möglicherweise zeichnen sich die acht *B. toyonensis* Isolate, die die genetische Information für alle neun Virulenzfaktoren (Nhe, Nhe<sub>a</sub>, HBL, HBL<sub>a</sub> und InhA1-5) tragen, im Vergleich zu den restlichen Stämmen aus Block 19 (Abbildung 14) durch ein gesteigertes Pathogenitätspotential aus. Beachtenswert erscheint angesichts dessen der *B. toyonensis* Typstamm, von dem als Probiotikum in der Nutztierwirtschaft Gebrauch gemacht wird [168] und der nachweislich als apathogen und nicht Toxin produzierend gilt [441]. Diesem fehlt sowohl *nheABC<sub>a</sub>* als auch *hb/CDA<sub>a</sub>* (Abbildung 14) und es ist wahrscheinlich, dass der Stamm aufgrund einer Frameshift-Mutation nicht zur Synthese eines funktionalen InhA3 Proteins befähigt ist.

Stamm-übergreifend betrachtet sind sich die Virulenzfaktorprofile der Arten *B. thuringiensis* (Teilcluster 20-A) und *B. cereus* s.s. (Teilcluster 20-B) sehr ähnlich (Tabelle 23). Der bedeutendste Unterschied liegt im Vorkommen des *hlyII* Genes. Während das Gen in lediglich neun von 67 *B. cereus* s.s. Isolaten (~ 13 %) anzutreffen ist, gibt es keinen *B. thuringiensis* Stamm im Datensatz, der das Toxinogen nicht besitzt. Unter mehreren getesteten Säugetier-Erythrozyten reagierten die roten Blutkörperchen des Menschen mit am sensitivsten gegenüber einer HlyII induzierten Lyse [442]. Eine Untersuchung mit 57 *B. cereus* s.l. Stämmen ergab, dass *hlyII* ausschließlich bei Stämmen mit pathogenem Potential (Verursacher von Lebensmittelerkrankungen oder nicht-gastrointestinalen Infektionen des Menschen) vorlag, obschon nur ~ 28 % der krankheitsverursachenden Isolate das Toxin kodierten [443]. Die Aktivität von HlyII leitet bei menschlichen Monozyten sowie Maus Makrophagen den programmierten Zelltod (Apoptose) ein und eine Deletion von *hlyII* führt im Mausmodell zur Virulenzminderung (Attenuation) des Erregerstammes *B. thuringiensis* 407 [325]. Dem Hämolyse wird vor allem eine Rolle bei opportunistischen Infektionen zugeordnet, wobei hypothesiert wird, dass unter Eisen- und Glukose-limitierten Bedingungen die Expression des Toxins zur Hämolyse von Erythrozyten wie auch zur Apoptose von Makrophagen führt, wodurch aufgrund der Freisetzung von Nährstoffen und Eisen die Proliferation, Dissemination und Persistenz des Erregers sichergestellt wird [314, 324]. Vorausgesetzt die Mehrheit der Isolate aus Teilcluster 20-A (Abbildung 14) ist zur Expression von *hlyII* im Stande, kann angesichts der ansonsten vergleichbaren Virulenzfaktorprofile der

beiden Spezies durchaus gemutmaßt werden, dass *B. thuringiensis* (Teilcluster 20-A) im Vergleich zu *B. cereus* s.s. (Teilcluster 20-B) der potentere opportunistische Infektionserreger ist. Dem Klassifikationsschema von Guinebretière *et al.* zufolge wird die *B. cereus* s.l. Gruppe in sieben große phylogenetische Gruppen (I-VII) eingeteilt, die jeweils noch aus mehreren Untergruppen bestehen können [252]. Dabei ist der *B. cereus* Typstamm Mitglied der Untergruppe IV-2, wohingegen der *B. thuringiensis* Typstamm der Untergruppe IV-3 zugeordnet wird. Gesamtheitlich betrachtet weisen beide Untergruppen eine moderate (IV-3) bzw. hohe (IV-2) zytotoxische Aktivität gegenüber Caco-2-Zellen auf [160, 254]. Allerdings konnten auch in Untergruppe IV-2 (*B. thuringiensis*) Stämme mit hoher zellschädigender Wirkung ausgemacht werden. Der Grad der Zytotoxizität eines Isolates ist grundsätzlich abhängig von der untersuchten Zelllinie [444, 445]. Beim Vergleich von zehn eukaryotischen Zelllinien reagierten Vero- und HUVEC-Zellen am anfälligsten auf die Toxin-Überstände des Nhe Referenzstammes *B. cereus* NVH 0075-95 [444]. Im Gegensatz dazu zeigte der Nhe/HBL Referenzstamm *B. cereus* F4430-73 die größten toxischen Effekte bei Hep-G2- und A549-Zellen. Die Daten einer neueren Studie belegen, dass von 15 *B. cereus* s.l. Stämmen, die der phylogenetischen Gruppe IV zuzurechnen sind, die sieben Isolate, die mit dem *B. thuringiensis* Typstamm ein Monophylum bilden, über eine weit höhere durchschnittliche Zytotoxizität bei HeLa-Zellen verfügen als die acht Stämme, die mit dem *B. cereus* Typstamm phylogenetisch assoziiert sind [255]. Ungeachtet der Frage, welche Spezies letztendlich über das größere zellschädigende Potential verfügt, sollte die Einstufung von *B. thuringiensis* als Risikogruppe 1 Organismus [446] nochmals einer umfassenden Überprüfung unterzogen werden. Immerhin sind hoch zytotoxische Vertreter auch in dieser Art zu finden [254, 255] und im Kontrast zu *B. cereus* s.s. (Risikogruppe 2 Prokaryot) besteht die Möglichkeit, dass es sich bei *B. thuringiensis* um den potentere opportunistischen Keim handelt.

HlyIII ist kaum erforscht und die Bedeutung des Toxins für die Virulenz von *B. cereus* s.l. ist mehr oder weniger unbekannt [314]. Im Zuge dieser Arbeit wurde die Existenz eines homologen Proteins (HlyIIIH) nachgewiesen (Abbildung 18B). Obwohl das putative Hämolyisin kein Spezies-spezifisches An-/Abwesenheitsmuster aufweist (Abbildung 14), ist sein Vorkommen charakteristisch für die Blöcke 5 bis 20, womit von den beschriebenen Arten einzig *B. "manliponensis"* und *B. cytotoxicus hlyIIIH*-defizient sind.

Die Genomospezies *B. anthracis* umfasst geringstenfalls sämtliche Stämme aus Teilcluster 16-C (Abbildung 12). Ob zudem die beiden anderen Teilcluster aus Block 16 der Art zuzurechnen sind (die Indices ANI<sub>b11</sub> und ANI<sub>m</sub> sagen eine Claque-Gruppe vorher), ist bezogen auf die Frage der Eignung der Anthrax-Toxin Gene als Spezies-definierende Marker ohne Bedeutung. Bereits in Teilcluster 16-C (Abbildung 14) besitzen 21 Stämme die genetische Information für das Anthrax-Toxin nicht, womit aus rein phylogenomischer Sicht nicht jedem Stamm der Genomospezies die Fähigkeit zum Auslösen der Milzbranderkrankung gegeben ist. Ähnlich verhält es sich bei *B. thuringiensis*. Bt-Toxine (Cry, Cyt, Vip und Sip) kommen in mindestens zwölf Stämmen der Genomospezies (Teilcluster 20-A) nicht vor und sind über Artgrenzen hinweg zum Beispiel auch bei *B. cereus* s.s. (Teilcluster 20-B) oder *B. mycoides* (Teilcluster 12-D) Isolaten vorhanden (Abbildung 16 und Abbildung 17). Ungeachtet dessen wurde die Aufteilung emetischer *B. cereus* s.l. in mehrere (phylogenetisch) voneinander unabhängige Abstammungslinien (Abbildung 14) auch in anderen Studien [225, 249] beobachtet. Dass die Präsenz/Absenz-Muster der drei bedeutenden Virulenzfaktoren (Anthrax-Toxin, Bt-Toxine und Cereulid-Synthetase) nicht mit der *B. cereus* s.l. Phylogenie korrelieren und damit als Spezies Marker ungeeignet sind, kann darauf zurückgeführt werden, dass die genetischen Determinanten der Toxine Teil des Mobilomes der *B. cereus* s.l. Gruppe und damit Gegenstand horizontalen Gentransfers sind [287, 447, 448].

In dieser Arbeit erfolgte die Suche nach Spezies-spezifischen Charakteristika vor allem in Hinsicht auf die An-/Abwesenheitsprofile der untersuchten Virulenzfaktoren. Ob Sequenzähnlichkeiten zur Diskriminierung von *B. cereus* s.l. Arten herangezogen werden können, war dahingegen nur in eingeschränktem Maße Gegenstand der hiesigen Untersuchungen. Nichtsdestotrotz konnte aufgedeckt werden, dass in jedem Stamm des *B. cereus* s.l. Datensatzes genau eine Kopie des *hlyA* Genes existiert (1:1 Orthologie) und dass sich die Genomospezies-Cluster (Blöcke/Teilcluster), die mittels der phylogenomischen Indices ANI<sub>b11</sub>, ANI<sub>m</sub> und gANI vorhergesagt wurden, bis auf wenige Ausnahmen in der *hlyA* Einzelgen Phylogenie widerspiegeln (Abbildung 19). Beide Beobachtungen machen *hlyA* zu einem potentiellen Marker für eine schnelle Charakterisierung von *B. cereus* s.l. Isolaten auf Spezies-Ebene vergleichbar mit der Verwendung des *panC* Genes zur Eingruppierung von Stämmen in die sieben großen phylogenetischen Gruppen nach Guinebretière *et al.* [252, 254]. Jedoch sollte beachtet werden, dass mittels *hlyA* eine Abgrenzung von *B. thuringiensis* (Teilcluster 20-A) zu *B. cereus* s.s. (Teilcluster 20-

B) nicht in jedem Fall möglich ist. In Übereinstimmung mit den Erkenntnissen einer anderen Studie [254] konnte das auf die Spezies *B. cytotoxicus* beschränkte Vorkommen von *cytK-1* bestätigt werden (Abbildung 18A). Darüber hinaus wurde enthüllt, dass die Art zudem über eine eigene Bikarbonat-Transporter Variante verfügt (Abbildung 15). So liegt die maximale Ähnlichkeit zwischen den *B. cytotoxicus* Substrat-Bindeproteinsequenzen und allen anderen *B. cereus s.l.* Substrat-Bindeproteinen bei 81,4 %, wohingegen alle anderen Substrat-Bindeproteinsequenzen untereinander eine Ähnlichkeit von mindestens 90,5 % aufweisen. Zusätzlich zum Substrat-Bindeprotein ist der ABC-Transporter noch aus zwei weiteren Proteinkomponenten (Permease und ATP-Bindeprotein) aufgebaut.

Auf Sequenzebene könnten Spezies-definierende Merkmale auch in Form von Sequenzsignaturen oder Polymorphismen vorhanden sein. Analog dazu ließ sich beispielsweise nachweisen, dass bei allen *B. cereus s.l.* Isolaten der phylogenetischen Gruppe VI (*sensu* Guinebretière *et al.*) die Sequenzmotive für Psychrotoleranz im Cold Shock Protein A (CspA) [449] und 16S rDNA Gen [450] vorliegen [252, 451]. Ein Screening nach Spezies-spezifischen Signaturen hätte den Untersuchungsrahmen dieser Studie gesprengt und ist deshalb nicht Teil der vorliegenden Arbeit. Allerdings wurden drei Einzelnukleotid-Polymorphismen gefunden, die charakteristisch für zwei monophyletische Gruppen (nicht Genomospezies!) sind. Die acht Stämme des klonalen *B. anthracis* Komplexes tragen eine Nonsense-Mutation im *nheC* Gen sowie eine Frameshift-Mutation im *hlyII* Gen. Beide Polymorphismen sollten jeweils ein trunkiertes Genprodukt zur Folge haben. Inwieweit die verkürzten Proteine ihre biologische Funktion noch ausüben können ist unklar. Neben der Nonsense-Mutation im *plcR* Gen [245] existieren somit noch zwei weitere Einzelnukleotid-Variationen, die spezifisch für den klonalen Komplex sind und wahrscheinlich einen Einfluss auf dessen pathogenes Verhalten haben. Das zweite Monophylum, das von den Anthrax-Toxin kodierenden Stämmen *B. cereus* 03BB87, *B. cereus* G9241 und *B. cereus* BcFL2013 gebildet wird (Teilcluster 16-A in Abbildung 14), ist durch eine Punktmutation im Startcodon (ATG → ATA) des *hblD* Genes gekennzeichnet, was zu einem verzögerten Translationsstart führen sollte. Auch hier ist aufgrund des zu erwartenden trunkierten Proteins dessen Funktionsverlust nicht ausgeschlossen. Inwiefern Spezies-charakterisierende Polymorphismen in Gestalt stiller oder nicht-synonymer Punktmutationen in der *B. cereus s.l.* Gruppe vorkommen, wurde nicht analysiert. Für die 28 Genomospezies-Cluster des hiesigen Datensatzes (Abbildung 23) konnten

jedenfalls keine Spezies-spezifischen Frameshift- oder Nonsense-Mutationen ausgemacht werden.

Abschließend ist festzuhalten, dass sich die untersuchten *B. cereus s.l.* Virulenzfaktoren allein auf Grundlage ihres An-/Abwesenheitsprofils nicht als Genomospezies-diskriminierende Marker eignen.

### 4.3. Das *B. cereus s.l.* Toxinovar System

Die Namensgebung von *B. cereus s.l.* Isolaten stimmt allzu oft nicht mit der tatsächlichen phylogenetischen/phylogenomischen Stellung der Stämme überein (Abbildung 23), was in vielen Fällen der historisch gewachsenen Benennung entsprechend phänotypischer Kriterien geschuldet ist. Gerade die Tatsache, dass es keine Virulenzfaktoren zu geben scheint, die einzig anhand ihrer bloßen An- bzw. Abwesenheit die Fähigkeit zur Abgrenzung von *B. cereus s.l.* Genomospezies auf sich vereinen (Abbildung 14), stellt die bisherige Praxis einer Spezieszuordnung, die vorrangig nach dem pathogenen Erscheinungsbild eines Isolates ausgerichtet ist, mehr als nur in Frage. Diskrepanzen, die sich in Zusammenhang mit der medizinischen Bedeutung und der phylogenetischen Stellung von Stämmen ergeben, werden in zunehmendem Maße deutlich. So existieren Milzbrand verursachende Stämme [241, 242, 244, 452], die nicht dem klonalen *B. anthracis* Komplex angehören (Abbildung 14). Zwar gibt es Vorschläge zur Benennung derartiger Borderline-Stämme als "*B. cereus* Varietät *anthracis*" [452] oder auch "*B. cereus/B. anthracis sensu lato*" [453], jedoch lösen derartige Ansätze die Problematik nur punktuell. Denn in ähnlicher Weise erstrecken sich derlei geartete Diskrepanzen zudem auf *B. cereus s.l.* Isolate mit agrarökonomischem Stellenwert (Abbildung 16 und Abbildung 17). Eine adäquate Lösung dieser Diskrepanzproblematik kann nur gelingen, wenn bei der Namensgebung eines beliebigen *B. cereus s.l.* Stammes die Brücke geschlagen werden kann zwischen dessen phylogenetischer Stellung, die Ausdruck in der Genomospezieszugehörigkeit findet, und seiner medizinischen/agrarökonomischen Bedeutung, die sich im Phänotyp niederschlägt. Eine solche Lösung, die im Stande ist beide taxonomisch relevanten Charakteristika (Phylogenie und Phänotyp) in einer sinnvollen Nomenklatur für die *B. cereus s.l.* Gruppe zu vereinen, wird durch das nachfolgend vorgeschlagene "Toxinovar System" repräsentiert.

Im Wesentlichen handelt es sich beim Toxinovar System um eine einziffrige Oktalnotation, die die medizinische/agrarökonomische Bedeutung eines *B. cereus s.l.* Stammes numerisch kodiert. Dabei werden den drei Virulenzfaktoren Anthrax-Toxin, Bt-Toxin und Cereulid-Synthetase die Oktalzahlen 1, 2 und 4 zugewiesen. Sofern ein *B. cereus s.l.* Stamm über keinen dieser drei Virulenzfaktoren verfügt, wird die Oktalzahl 0 vergeben. Die Namensgebung erfolgt wie üblich binomial, wobei der Arname strikt die Genomospezieszugehörigkeit des zu benennenden Isolates (und somit dessen phylogenetische Stellung) ausdrückt.

#### Das Toxinovar 0 (Präsenz von Diarrhoe Toxinen)

Als einführendes Beispiel zur Erläuterung der Toxinovar Nomenklatur soll der Stamm *B. mycoides* DSM 307 dienen. Da der Stamm in Block 9 ansässig ist, gehört er der Genomospezies *B. pseudomycoides* an (Abbildung 12). Darüber hinaus besitzt das Isolat keinen der obig aufgeführten Virulenzfaktoren (Abbildung 14 und Abbildung 16), was numerisch durch eine 0 ausgedrückt wird. Nach dem Toxinovar System erhält der Stamm dementsprechend die Bezeichnung *B. pseudomycoides* DSM 307 Toxinovar 0. Zu beachten ist, dass ein Stamm mit Zugehörigkeit zum Toxinovar 0 keinesfalls medizinisch unbedeutend sein muss. Vielmehr verhält es sich so, dass durch diese Oktalzahl eine Art Grundfähigkeit der *B. cereus s.l.* Gruppe zum Ausdruck kommt. Diese entspricht der Befähigung zum Auslösen des Diarrhoe Syndroms, wozu (rein theoretisch!) jedes *B. cereus s.l.* Isolat zu einem gewissen Grad im Stande sein sollte, da es so gut wie keinen Stamm gibt, der nicht zumindest über einen der drei Hauptvirulenzfaktoren dieser Lebensmittelerkrankung (Nhe, HBL, CytK) verfügt (Abbildung 14).

#### Die Toxinovare 1 und 1\* (Milzbrand verursachende Stämme)

Aufbauend auf den Erläuterungen des vorausgegangenen Abschnitts wird nun der Stamm *B. cereus* 03BB102 betrachtet. Dieser kodiert das Anthrax-Toxin (Abbildung 14) und ist daher als (potentieller) Milzbranderreger zu werten. Dem Virulenzfaktor (Anthrax-Toxin) wird die Oktalzahl 1 zugewiesen. Da das betrachtete Isolat Mitglied des Teilclusters 16-C ist, wird es seiner Genomospezieszugehörigkeit entsprechend mit *B. anthracis* 03BB102 Toxinovar 1 benannt. An dieser Stelle ist hervorzuheben, dass in der neuen Bezeichnung eine Entkoppelung des medizinischen Phänotyps (Milzbrandkrankung) vom Artnamen stattgefunden hat. Nicht mehr der Arname (*B. anthracis*) weist daraufhin, dass es sich um einen Milzbranderreger handelt, sondern einzig das Toxinovar 1.

Sofern es sich beim zu benennenden Isolat um einen Vertreter des klonalen *B. anthracis* Komplexes handeln sollte, kommt es zusätzlich zur Kennzeichnung mit einem “\*“-Symbol. Demzufolge wird eine Bezeichnung wie *B. anthracis* Vollum abgelöst von *B. anthracis* Vollum Toxinovar 1\*.

#### Das Toxinovar 2 (entomopathogene *B. cereus s.l.*)

*B. cereus s.l.* Stämme werden dem Toxinovar 2 zugeordnet, falls sie über die genetische Information für mindestens ein Bt-Toxin (Cry, Cyt, Vip oder Sip) verfügen. Damit trägt dieses Toxinovar der agrarökonomischen Bedeutung von *B. cereus s.l.* Rechnung und bezieht sich auf Isolate mit möglicher insektizider Wirkung. Als Fallbeispiel wird das Isolat *B. thuringiensis* BGSC 4AJ1, das zwei Bt-Toxin Gene trägt und der Genomospezies *B. anthracis* (Teilcluster 16-C) angehört (Abbildung 16), herangezogen. Konform zur neuen Nomenklatur wird der Stamm mit *B. anthracis* BGSC 4AJ1 Toxinovar 2 bezeichnet. Auch hier sollte verinnerlicht werden, dass auf Grundlage des neuen Artnamens kein Rückschluss mehr auf den pathogenen Phänotyp gezogen werden darf. Das Isolat besitzt weder das Anthrax-Toxin noch ist dessen Speziesname “*thuringiensis*“. Allein die Oktalzahl 2 zeigt an, dass der Stamm mit Bt-Toxin Genen ausgestattet ist.

Wenn ein Stamm ausschließlich putative Bt-Toxine kodiert, folgt der Oktalzahl der Buchstabe “p“ (für putativ). Nach dem Regelwerk der *B. thuringiensis* Toxin Nomenklatur gilt ein Bt-Toxin als putativ, sofern es < 95 % Sequenzidentität (auf Aminosäureebene) zur ähnlichsten bereits verifizierten Toxinsequenz aufweist (<http://www.btnomenclature.info/>).

#### Das Toxinovar 4 (emetische *B. cereus s.l.*)

Das Toxinovar 4 ist emetischen *B. cereus s.l.* vorbehalten und somit denjenigen Stämmen, die das Cereulid-Synthetase-Gencluster enthalten. So wird beispielsweise der Stamm *B. cereus* BtB2-4 zu *B. mycoides* BtB2-4 Toxinovar 4 umbenannt. Klar ersichtlich ist, dass im Gegensatz zur ursprünglichen Namensgebung aus dem neuen Bezeichner sowohl die phylogenetische Stellung des Stammes (Mitglied der Genomospezies *B. mycoides* aus Teilcluster 12-D) wie auch seine Kompetenz zur Produktion der Cereulid-Synthetase hervorgeht (Abbildung 14).

#### Die bislang hypothetischen Toxinovare 3, 5, 6 und 7

Neben der strikten Trennung von Phylogenie und pathogenem Phänotyp bei der Benennung von Isolaten, liegt eine der Stärken des Toxinovar Systems in seiner

Erweiterbarkeit auf Kombinationen von Virulenzfaktoren, die in der *B. cereus s.l.* Gruppe bislang (noch) nicht beobachtet wurden, jedoch nicht auszuschließen sind. Im Folgenden wird von einem hypothetischen Stamm ausgegangen, der die Bezeichnung *B. cereus* Hyp1 Toxinovar 6 trägt. Demnach gehört das Isolat zur Genomospezies *B. cereus s.s.* und weil sich das Toxinovar 6 aus den beiden grundlegenden Toxinovaren 2 und 4 ( $2 + 4 = 6$ ) ableitet, verfügt der Stamm über mindestens ein Bt-Toxin Gen (Toxinovar 2) sowie das Cereulid-Synthetase-Gencluster (Toxinovar 4). Sämtliche Toxinovare, die sich mit Hilfe der Oktalnotation realisieren lassen, sind in Tabelle 25 zusammengefasst.

**Tabelle 25 Grundlegende und ableitbare Toxinovare**

**Grundlegende Toxinovare**

Oktalzahl	vorhandene Virulenzfaktoren (Gene/Gencluster)
1 (bzw. 1*)	Anthrax-Toxin ( <i>cya</i> , <i>pagA</i> und <i>lef</i> )
2	Bt-Toxin(e) ( <i>cry</i> , <i>cyt</i> , <i>vip</i> und/oder <i>sip</i> )
4	Cereulid-Synthetase ( <i>cesHPTABCD</i> )

**Ableitbare Toxinovare**

Oktalzahl	vorhandene Virulenzfaktoren
3 (= 1+2)	Anthrax-Toxin & Bt-Toxin(e)
5 (= 1+4)	Anthrax-Toxin(e) & Cereulid-Synthetase
6 (= 2+4)	Bt-Toxin(e) & Cereulid-Synthetase
7 (= 1+2+4)	Anthrax-Toxin & Bt-Toxin(e) & Cereulid-Synthetase

Genomospezies ohne Artnamen/Typstamm

Für den Fall, dass der zu bezeichnende Stamm einer Genomospezies ohne Artnamen angehört (kein beschriebener Typstamm im Block bzw. Teilcluster), wird in der Toxinovar Nomenklatur solange der Platzhalter "*B. cereus* clade" verwendet bis es zur Neubeschreibung einer entsprechenden Nomenspezies gekommen ist.

In Tabelle 26 sind alle bisher beschriebenen Beispiele zur Namensgebung nach den Regeln der *B. cereus s.l.* Toxinovar Nomenklatur aufgeführt. Darüber hinaus beinhaltet die Tabelle weitere Fallbeispiele zur Verdeutlichung der dem System innewohnenden Logik.

Tabelle 26 Fallbeispiele zum Toxinovar System

## Erläuterte Beispiele

NCBI Nomenklatur	Genomospezieszugehörigkeit (Block/Teilcluster in Abbildung 14)	Stamm besitzt genetische Information für			Toxinovar Nomenklatur
		Anthrax-Toxin	Bt-Toxin(e)	Cereulid-Synthetase	
<i>B. mycooides</i> DSM 307	<i>B. pseudomycooides</i> (Block 9)				<i>B. pseudomycooides</i> DSM 307 Toxinovar 0
<i>B. cereus</i> 03BB102	<i>B. anthracis</i> (Teilcluster 16-C)	x			<i>B. anthracis</i> 03BB102 Toxinovar 1
<i>B. anthracis</i> Vollum	Klonaler <i>B. anthracis</i> Komplex (Teilcluster 16-C)	x			<i>B. anthracis</i> Vollum Toxinovar 1*
<i>B. thuringiensis</i> BGSC 4AJ1	<i>B. anthracis</i> (Teilcluster 16-C)		x		<i>B. anthracis</i> BGSC 4AJ1 Toxinovar 2
<i>B. cereus</i> BtB2-4	<i>B. mycooides</i> (Teilcluster 12-D)			x	<i>B. mycooides</i> BtB2-4 Toxinovar 4
hypothetisches Isolat	<i>B. cereus</i> s.s.		x	x	<i>B. cereus</i> Hyp1 Toxinovar 6

## Weitere Beispiele

NCBI Nomenklatur	Genomospezieszugehörigkeit (Block/Teilcluster in Abbildung 14)	Stamm besitzt genetische Information für			Toxinovar Nomenklatur
		Anthrax-Toxin	Bt-Toxin(e)	Cereulid-Synthetase	
<i>B. "gaemokensis"</i> JCM 15801 <sup>T</sup>	<i>B. "gaemokensis"</i> (Block 5)		x (nur putative Bt-Toxine)		<i>B. "gaemokensis"</i> JCM 15801 <sup>T</sup> Toxinovar 2p
<i>B. sp.</i> UNC437CL72CviS29	kein Arname (Block 8)				<i>B. cereus</i> clade UNC437CL72CviS29 Toxinovar 0
<i>B. cereus</i> MC67	kein Arname (Teilcluster 12-C)			x	<i>B. cereus</i> clade MC67 Toxinovar 4
<i>B. cereus</i> HuA2-1	<i>B. mycooides</i> (Teilcluster 12-D)		x (Nonsense-Mutation)		<i>B. mycooides</i> HuA2-1 Toxinovar 2
<i>B. cereus</i> G9241	kein Arname (Teilcluster 16-A)	x			<i>B. cereus</i> clade G9241 Toxinovar 1
<i>B. cereus</i> BDRD-ST26	kein Arname (Teilcluster 16-A)			x (Frameshift-Mutation)	<i>B. cereus</i> clade BDRD-ST26 Toxinovar 4
<i>B. thuringiensis</i> MC28	<i>B. toyonensis</i> (Block 19)		x		<i>B. toyonensis</i> MC28 Toxinovar 2
<i>B. cereus</i> MSX-A1	<i>B. thuringiensis</i> (Teilcluster 20-A)				<i>B. thuringiensis</i> MSX-A1 Toxinovar 0
<i>B. thuringiensis</i> AK47	<i>B. cereus</i> s.s. (Teilcluster 20-B)		x		<i>B. cereus</i> AK47 Toxinovar 2
hypothetisches Isolat	<i>B. mycooides</i>	x	x		<i>B. mycooides</i> Hyp2 Toxinovar 3
hypothetisches Isolat	kein Arname	x		x	<i>B. cereus</i> clade Hyp3 Toxinovar 5
hypothetisches Isolat	Klonaler <i>B. anthracis</i> Komplex	x	x (nur putative Bt-Toxine)	x	<i>B. anthracis</i> Hyp4 Toxinovar 7p*

## VII. LITERATURVERZEICHNIS

1. Rossello-Mora R, Amann R: **The species concept for prokaryotes.** *FEMS Microbiol Rev* 2001, **25**(1):39-67.
2. Stackebrandt E, Frederiksen W, Garrity GM, Grimont PA, Kämpfer P, Maiden MC, Nesme X, Rosselló-Mora R, Swings J, Trüper HG: **Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology.** *International journal of systematic and evolutionary microbiology* 2002, **52**(3):1043-1047.
3. Rossello-Mora R, Amann R: **Past and future species definitions for Bacteria and Archaea.** *Syst Appl Microbiol* 2015, **38**(4):209-216.
4. Vandamme P, Pot B, Gillis M, de Vos P, Kersters K, Swings J: **Polyphasic taxonomy, a consensus approach to bacterial systematics.** *Microbiol Rev* 1996, **60**(2):407-438.
5. Tindall BJ, Rossello-Mora R, Busse HJ, Ludwig W, Kämpfer P: **Notes on the characterization of prokaryote strains for taxonomic purposes.** *Int J Syst Evol Microbiol* 2010, **60**(Pt 1):249-266.
6. Brenner DJ, Fanning GR, Rake AV, Johnson KE: **Batch procedure for thermal elution of DNA from hydroxyapatite.** *Analytical biochemistry* 1969, **28**:447-459.
7. Brenner DJ: **Deoxyribonucleic acid reassociation in the taxonomy of enteric bacteria.** *International Journal of Systematic and Evolutionary Microbiology* 1973, **23**(4):298-307.
8. Johnson JL: **Use of nucleic-acid homologies in the taxonomy of anaerobic bacteria.** *International Journal of Systematic and Evolutionary Microbiology* 1973, **23**(4):308-315.
9. Wayne LG, Brenner DJ, Colwell RR, Grimont PA, Kandler O, Krichevsky MI, Moore LH, Moore WE, Murray RG, Stackebrandt E: **Report of the ad hoc committee on reconciliation of approaches to bacterial systematics.** *International Journal of Systematic and Evolutionary Microbiology* 1987, **37**(4):463-464.
10. Fox GE, Pechman KR, Woese CR: **Comparative cataloging of 16S ribosomal ribonucleic acid: molecular approach to procaryotic systematics.** *International Journal of Systematic and Evolutionary Microbiology* 1977, **27**(1):44-57.
11. Woese CR, Fox GE: **Phylogenetic structure of the prokaryotic domain: the primary kingdoms.** *Proc Natl Acad Sci U S A* 1977, **74**(11):5088-5090.
12. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, Rooney AP, Yi H, Xu XW, De Meyer S *et al.*: **Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes.** *Int J Syst Evol Microbiol* 2018, **68**(1):461-466.
13. Sanger F, Nicklen S, Coulson AR: **DNA sequencing with chain-terminating inhibitors.** *Proc Natl Acad Sci U S A* 1977, **74**(12):5463-5467.
14. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA: **Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase.** *Science* 1988, **239**(4839):487-491.
15. Ludwig W, Schleifer KH: **Bacterial phylogeny based on 16S and 23S rRNA sequence analysis.** *FEMS Microbiol Rev* 1994, **15**(2-3):155-173.
16. Stackebrandt E, Goebel BM: **Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology.** *International Journal of Systematic and Evolutionary Microbiology* 1994, **44**(4):846-849.
17. Stackebrandt E, Ebers J: **Taxonomic parameters revisited: tarnished gold standards.** *Microbiol Today* 2006, **33**:152-155.
18. Kämpfer P, Glaeser SP: **Prokaryotic taxonomy in the sequencing era--the polyphasic approach revisited.** *Environ Microbiol* 2012, **14**(2):291-317.
19. Glaeser SP, Kämpfer P: **Multilocus sequence analysis (MLSA) in prokaryotic taxonomy.** *Syst Appl Microbiol* 2015, **38**(4):237-245.
20. Zeigler DR: **Gene sequences useful for predicting relatedness of whole genomes in bacteria.** *Int J Syst Evol Microbiol* 2003, **53**(Pt 6):1893-1900.
21. Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, Stackebrandt E, Van de Peer Y, Vandamme P, Thompson FL *et al.*: **Opinion: Re-evaluating prokaryotic species.** *Nat Rev Microbiol* 2005, **3**(9):733-739.

22. Schleifer KH: **Classification of Bacteria and Archaea: past, present and future.** *Syst Appl Microbiol* 2009, **32**(8):533-542.
23. Guo Y, Zheng W, Rong X, Huang Y: **A multilocus phylogeny of the *Streptomyces griseus* 16S rRNA gene clade: use of multilocus sequence analysis for streptomycete systematics.** *Int J Syst Evol Microbiol* 2008, **58**(Pt 1):149-159.
24. Vanlaere E, Baldwin A, Gevers D, Henry D, De Brandt E, LiPuma JJ, Mahenthiralingam E, Speert DP, Dowson C, Vandamme P: **Taxon K, a complex within the *Burkholderia cepacia* complex, comprises at least two novel species, *Burkholderia contaminans* sp. nov. and *Burkholderia lata* sp. nov.** *Int J Syst Evol Microbiol* 2009, **59**(Pt 1):102-111.
25. Mulet M, Lalucat J, Garcia-Valdes E: **DNA sequence-based analysis of the *Pseudomonas* species.** *Environ Microbiol* 2010, **12**(6):1513-1530.
26. Bohm ME, Huptas C, Krey VM, Scherer S: **Massive horizontal gene transfer, strictly vertical inheritance and ancient duplications differentially shape the evolution of *Bacillus cereus* enterotoxin operons hbl, cytK and nhe.** *BMC Evol Biol* 2015, **15**:246.
27. Rossello-Mora R: **Towards a taxonomy of Bacteria and Archaea based on interactive and cumulative data repositories.** *Environ Microbiol* 2012, **14**(2):318-334.
28. Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi M, Sentausa E, Raoult D, Fournier PE: **A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species.** *Int J Syst Evol Microbiol* 2014, **64**(Pt 2):384-391.
29. Oren A, Garrity GM: **Then and now: a systematic review of the systematics of prokaryotes in the last 80 years.** *Antonie Van Leeuwenhoek* 2014, **106**(1):43-56.
30. Karlsson R, Gonzales-Siles L, Boulund F, Svensson-Stadler L, Skovbjerg S, Karlsson A, Davidson M, Hulth S, Kristiansson E, Moore ER: **Proteotyping: Proteomic characterization, classification and identification of microorganisms--A prospectus.** *Syst Appl Microbiol* 2015, **38**(4):246-257.
31. Welker M, Moore ER: **Applications of whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry in systematic microbiology.** *Syst Appl Microbiol* 2011, **34**(1):2-11.
32. Wenning M, Scherer S: **Identification of microorganisms by FTIR spectroscopy: perspectives and limitations of the method.** *Appl Microbiol Biotechnol* 2013, **97**(16):7111-7120.
33. Krieg NR: **Bacterial classification: an overview.** *Can J Microbiol* 1988, **34**(4):536-540.
34. Rosselló-Mora R: **DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation.** In: *Molecular identification, systematics, and population structure of prokaryotes.* Springer; 2006: 23-50.
35. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM *et al.* **Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.** *Science* 1995, **269**(5223):496-512.
36. Chun J, Rainey FA: **Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea.** *Int J Syst Evol Microbiol* 2014, **64**(Pt 2):316-324.
37. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C: **Ten years of next-generation sequencing technology.** *Trends Genet* 2014, **30**(9):418-426.
38. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z *et al.* **Genome sequencing in microfabricated high-density picolitre reactors.** *Nature* 2005, **437**(7057):376-380.
39. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR *et al.* **Accurate whole human genome sequencing using reversible terminator chemistry.** *Nature* 2008, **456**(7218):53-59.
40. Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek JA, Costa G, McKernan K *et al.* **A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning.** *Genome research* 2008, **18**(7):1051-1063.
41. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H: **Continuous base identification for single-molecule nanopore DNA sequencing.** *Nat Nanotechnol* 2009, **4**(4):265-270.
42. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B *et al.* **Real-time DNA sequencing from single polymerase molecules.** *Science* 2009, **323**(5910):133-138.

43. Park ST, Kim J: **Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing.** *Int Neurol J* 2016, **20**(Suppl 2):S76-83.
44. Reuter JA, Spacek DV, Snyder MP: **High-throughput sequencing technologies.** *Mol Cell* 2015, **58**(4):586-597.
45. Goodwin S, McPherson JD, McCombie WR: **Coming of age: ten years of next-generation sequencing technologies.** *Nature reviews Genetics* 2016, **17**(6):333-351.
46. Konstantinidis KT, Tiedje JM: **Genomic insights that advance the species definition for prokaryotes.** *Proc Natl Acad Sci U S A* 2005, **102**(7):2567-2572.
47. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM: **DNA-DNA hybridization values and their relationship to whole-genome sequence similarities.** *Int J Syst Evol Microbiol* 2007, **57**(Pt 1):81-91.
48. Richter M, Rossello-Mora R: **Shifting the genomic gold standard for the prokaryotic species definition.** *Proc Natl Acad Sci U S A* 2009, **106**(45):19126-19131.
49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**(3):403-410.
50. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL: **Versatile and open software for comparing large genomes.** *Genome Biol* 2004, **5**(2):R12.
51. Richter M, Rossello-Mora R, Glockner OF, Peplies J: **JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison.** *Bioinformatics* 2016, **32**(6):929-931.
52. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A: **Microbial species delineation using whole genome sequences.** *Nucleic Acids Res* 2015, **43**(14):6761-6771.
53. Lee I, Ouk Kim Y, Park SC, Chun J: **OrthoANI: An improved algorithm and software for calculating average nucleotide identity.** *Int J Syst Evol Microbiol* 2016, **66**(2):1100-1103.
54. Yoon SH, Ha SM, Lim J, Kwon S, Chun J: **A large-scale evaluation of algorithms to calculate average nucleotide identity.** *Antonie Van Leeuwenhoek* 2017, **110**(10):1281-1286.
55. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S: **High-throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries.** *bioRxiv* 2017:225342.
56. Henz SR, Huson DH, Auch AF, Nieselt-Struwe K, Schuster SC: **Whole-genome prokaryotic phylogeny.** *Bioinformatics* 2005, **21**(10):2329-2335.
57. Auch AF, Henz SR, Holland BR, Goker M: **Genome BLAST distance phylogenies inferred from whole plastid and whole mitochondrion genome sequences.** *BMC Bioinformatics* 2006, **7**:350.
58. Auch AF, von Jan M, Klenk HP, Goker M: **Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison.** *Stand Genomic Sci* 2010, **2**(1):117-134.
59. Meier-Kolthoff JP, Auch AF, Klenk H-P, Goker M: **Genome sequence-based species delimitation with confidence intervals and improved distance functions.** *BMC Bioinformatics* 2013, **14**:60.
60. Sutcliffe IC, Trujillo ME, Goodfellow M: **A call to arms for systematists: revitalising the purpose and practises underpinning the description of novel microbial taxa.** *Antonie Van Leeuwenhoek* 2012, **101**(1):13-20.
61. Vandamme P, Peeters C: **Time to revisit polyphasic taxonomy.** *Antonie Van Leeuwenhoek* 2014, **106**(1):57-65.
62. Thompson CC, Amaral GR, Campeao M, Edwards RA, Polz MF, Dutilh BE, Ussery DW, Sawabe T, Swings J, Thompson FL: **Microbial taxonomy in the post-genomic era: rebuilding from scratch?** *Arch Microbiol* 2015, **197**(3):359-370.
63. Christensen H, Bisgaard M, Frederiksen W, Muttters R, Kuhnert P, Olsen JE: **Is characterization of a single isolate sufficient for valid publication of a new genus or species? Proposal to modify recommendation 30b of the Bacteriological Code (1990 Revision).** *Int J Syst Evol Microbiol* 2001, **51**(Pt 6):2221-2225.
64. Felis GE, Dellaglio F: **On species descriptions based on a single strain: proposal to introduce the status species proponenda (sp. pr.).** *International journal of systematic and evolutionary microbiology* 2007, **57**(9):2185-2187.

65. Forde BM, O'Toole PW: **Next-generation sequencing technologies and their impact on microbial genomics.** *Brief Funct Genomics* 2013, **12**(5):440-453.
66. Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Verezemskaja O, Isbandi M, Thomas AD, Ali R, Sharma K, Kyripides NC *et al*: **Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements.** *Nucleic Acids Res* 2017, **45**(D1):D446-D456.
67. Parte AC: **LPSN--list of prokaryotic names with standing in nomenclature.** *Nucleic Acids Res* 2014, **42**(Database issue):D613-616.
68. Euzéby JP: **List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet.** *Int J Syst Bacteriol* 1997, **47**(2):590-592.
69. Cowan ST: **Heretical taxonomy for bacteriologists.** *J Gen Microbiol* 1970, **61**(2):145-154.
70. Staley JT: **Comprehending microbial diversity: the fourth goal of microbial taxonomy.** *The Bulletin of BISMIS* 2010, **1**(1):1-5.
71. Curtis TP, Sloan WT, Scannell JW: **Estimating prokaryotic diversity and its limits.** *Proc Natl Acad Sci U S A* 2002, **99**(16):10494-10499.
72. Pedros-Alio C: **Marine microbial diversity: can it be determined?** *Trends Microbiol* 2006, **14**(6):257-263.
73. Lapage SP, Sneath PHA, Lessel EF, Skerman VBD, Seeliger HPR, Clark WA: **International code of nomenclature of bacteria (1990 revision):** American Society for Microbiology; 1992.
74. Vartoukian SR, Palmer RM, Wade WG: **Strategies for culture of 'unculturable' bacteria.** *FEMS Microbiol Lett* 2010, **309**(1):1-7.
75. Puspita ID, Kamagata Y, Tanaka M, Asano K, Nakatsu CH: **Are uncultivated bacteria really uncultivable?** *Microbes Environ* 2012, **27**(4):356-366.
76. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, Darling A, Malfatti S, Swan BK, Gies EA *et al*: **Insights into the phylogeny and coding potential of microbial dark matter.** *Nature* 2013, **499**(7459):431-437.
77. Gawad C, Koh W, Quake SR: **Single-cell genome sequencing: current state of the science.** *Nature reviews Genetics* 2016, **17**(3):175-188.
78. Whitman WB: **Genome sequences as the type material for taxonomic descriptions of prokaryotes.** *Syst Appl Microbiol* 2015, **38**(4):217-222.
79. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ *et al*: **A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea.** *Nature* 2009, **462**(7276):1056-1060.
80. Federhen S: **Type material in the NCBI Taxonomy Database.** *Nucleic Acids Res* 2015, **43**(Database issue):D1086-1098.
81. Sutcliffe IC: **A phylum level perspective on bacterial cell envelope architecture.** *Trends Microbiol* 2010, **18**(10):464-470.
82. Amaral GR, Dias GM, Wellington-Oguri M, Chimetto L, Campeao ME, Thompson FL, Thompson CC: **Genotype to phenotype: identification of diagnostic vibrio phenotypes using whole genome sequences.** *Int J Syst Evol Microbiol* 2014, **64**(Pt 2):357-365.
83. Mardis ER: **The impact of next-generation sequencing technology on genetics.** *Trends Genet* 2008, **24**(3):133-141.
84. **Global NGS market: Most lucrative sector of the genomics industry** [<http://www.companiesandmarkets.com/Market/Healthcare-and-Medical/Market-Research/Next-Generation-Sequencing-NGS-Market-Global-Forecast-to-2017/RPT1167432>]
85. Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, Swerdlow H, Turner DJ: **A large genome center's improvements to the Illumina sequencing system.** *Nat Methods* 2008, **5**(12):1005-1010.
86. Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ: **Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes.** *Nat Methods* 2009, **6**(4):291-295.
87. Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C, Gnirke A: **Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries.** *Genome Biol* 2011, **12**(2):R18.
88. Oyola SO, Otto TD, Gu Y, Maslen G, Manske M, Campino S, Turner DJ, Macinnis B, Kwiatkowski DP, Swerdlow HP *et al*: **Optimizing Illumina next-generation sequencing library preparation for extremely AT-biased genomes.** *BMC Genomics* 2012, **13**:1.

89. Tyler AD, Christianson S, Knox NC, Mabon P, Wolfe J, Van Domselaar G, Graham MR, Sharma MK: **Comparison of Sample Preparation Methods Used for the Next-Generation Sequencing of Mycobacterium tuberculosis.** *PloS one* 2016, **11**(2):e0148676.
90. Schatz MC, Delcher AL, Salzberg SL: **Assembly of large genomes using second-generation sequencing.** *Genome research* 2010, **20**(9):1165-1173.
91. O'Brien HE, Gong Y, Fung P, Wang PW, Guttman DS: **Use of low-coverage, large-insert, short-read data for rapid and accurate generation of enhanced-quality draft Pseudomonas genome sequences.** *PloS one* 2011, **6**(11):e27199.
92. Magoc T, Pabinger S, Canzar S, Liu X, Su Q, Puiu D, Tallon LJ, Salzberg SL: **GAGE-B: an evaluation of genome assemblers for bacterial organisms.** *Bioinformatics* 2013, **29**(14):1718-1725.
93. Lee H, Tang H: **Next-generation sequencing technologies and fragment assembly algorithms.** *Methods Mol Biol* 2012, **855**:155-174.
94. Salzberg SL, Phillippy AM, Zimin A, Puiu D, Magoc T, Koren S, Treangen TJ, Schatz MC, Delcher AL, Roberts M *et al*: **GAGE: A critical evaluation of genome assemblies and assembly algorithms.** *Genome research* 2012, **22**(3):557-567.
95. Haridas S, Breuill C, Bohlmann J, Hsiang T: **A biologist's guide to de novo genome assembly using next-generation sequence data: A test with fungal genomes.** *J Microbiol Methods* 2011, **86**(3):368-375.
96. Lin Y, Li J, Shen H, Zhang L, Papasian CJ, Deng HW: **Comparative studies of de novo assembly tools for next-generation sequencing technologies.** *Bioinformatics* 2011, **27**(15):2031-2037.
97. Desai A, Marwah VS, Yadav A, Jha V, Dhaygude K, Bangar U, Kulkarni V, Jere A: **Identification of optimum sequencing depth especially for de novo genome assembly of small genomes using next generation sequencing data.** *PloS one* 2013, **8**(4):e60204.
98. Earl D, Bradnam K, St John J, Darling A, Lin D, Fass J, Yu HO, Buffalo V, Zerbino DR, Diekhans M *et al*: **Assemblathon 1: a competitive assessment of de novo short read assembly methods.** *Genome research* 2011, **21**(12):2224-2241.
99. Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, Boisvert S, Chapman JA, Chapuis G, Chikhi R *et al*: **Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species.** *Gigascience* 2013, **2**(1):10.
100. Migula W: **Über ein neues System der Bakterien.** *Arb Bakteriell Inst Technisch Hochsch Karlsruhe* 1894(1):235-238.
101. Migula W: **System der Bakterien, Handbuch der Morphologie, Entwicklungsgeschichte und Systematik der Bakterien.** *Gustav Fisher, Jena, Germany* 1900, **Vol. 2.**
102. Palleroni NJ: **Genus I. Pseudomonas.** In: *Bergey's Manual of Systematic Bacteriology, Volume 2, The Proteobacteria, Part B: The Gammaproteobacteria.* Edited by Garrity GM, Brenner DJ, Krieg NR, Staley JT, vol. 2: Springer, New York; 2005: 323-379.
103. Kersters K, Ludwig W, Vancanneyt M, De Vos P, Gillis M, Schleifer K-H: **Recent changes in the classification of the pseudomonads: an overview.** *Systematic and Applied Microbiology* 1996, **19**(4):465-477.
104. Peix A, Ramirez-Bahena MH, Velazquez E: **Historical evolution and current status of the taxonomy of genus Pseudomonas.** *Infect Genet Evol* 2009, **9**(6):1132-1147.
105. Peix A, Ramirez-Bahena MH, Velazquez E: **The current status on the taxonomy of Pseudomonas revisited: An update.** *Infect Genet Evol* 2018, **57**:106-116.
106. Palleroni NJ, Kunisawa R, Contopoulou R, Doudoroff M: **Nucleic acid homologies in the genus Pseudomonas.** *International Journal of Systematic and Evolutionary Microbiology* 1973, **23**(4):333-339.
107. Palleroni NJ: **Genus I. Pseudomonas.** In: *Bergey's Manual of Systematic Bacteriology.* Edited by Krieg NR, Holt JG, vol. 1: Williams & Wilkins Co, Baltimore; 1984: 141-199.
108. De Vos P, De Ley J: **Intra-and intergeneric similarities of Pseudomonas and Xanthomonas ribosomal ribonucleic acid cistrons.** *International Journal of Systematic and Evolutionary Microbiology* 1983, **33**(3):487-509.
109. De Vos P, Goor M, Gillis M, De Ley J: **Ribosomal ribonucleic acid cistron similarities of phytopathogenic Pseudomonas species.** *International Journal of Systematic and Evolutionary Microbiology* 1985, **35**(2):169-184.

110. De Vos P, Van Landschoot A, Segers P, Tytgat R, Gillis M, Bauwens M, Rossau R, Goor M, Pot B, Kersters K: **Genotypic relationships and taxonomic localization of unclassified Pseudomonas and Pseudomonas-like strains by deoxyribonucleic acid: ribosomal ribonucleic acid hybridizations.** *International Journal of Systematic and Evolutionary Microbiology* 1989, **39**(1):35-49.
111. Schroeter J: **Ueber einige durch Bakterien gebildete Pigmente.** In: *Beiträge zur Biologie der Pflanzen.* Edited by Cohn F, vol. 1: J.U. Kern's Verlag, Breslau; 1872: 109-126.
112. Hugh R, Leifson E: **The proposed neotype strains of Pseudomonas aeruginosa (Schroeter 1872) Migula 1900.** *International Journal of Systematic and Evolutionary Microbiology* 1964, **14**(2):69-84.
113. Moore ERB, Tindall BJ, Dos Santos VAPM, Pieper DH, Ramos J-L, Palleroni NJ: **Nonmedical: Pseudomonas.** In: *The Prokaryotes, A Handbook on the Biology of Bacteria, Third Edition, Volume 6: Proteobacteria: Gamma Subclass.* Edited by Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E: Springer, New York; 2006: 646-703.
114. Woese CR, Weisburg WG, Hahn CM, Paster BJ, Zablén LB, Lewis BJ, Macke TJ, Ludwig W, Stackebrandt E: **The phylogeny of purple bacteria: the gamma subdivision.** *Systematic and Applied Microbiology* 1985, **6**(1):25-33.
115. Woese CR: **Bacterial evolution.** *Microbiol Rev* 1987, **51**(2):221-271.
116. Stackebrandt E, Murray RGE, Trüper HG: **Proteobacteria classis nov., a name for the phylogenetic taxon that includes the "purple bacteria and their relatives".** *International Journal of Systematic and Evolutionary Microbiology* 1988, **38**(3):321-325.
117. Garrity GM, Bell JA, Lilburn T: **Class III. Gammaproteobacteria.** In: *Bergey's Manual of Systematic Bacteriology, Volume 2, The Proteobacteria, Part B: The Gammaproteobacteria.* Edited by Garrity GM, Brenner DJ, Krieg NR, Staley JT, vol. 2: Springer, New York; 2005: 1.
118. Moore ERB, Mau M, Arnscheidt A, Böttger EC, Hutson RA, Collins MD, Van De Peer Y, De Wachter R, Timmis KN: **The Determination and Comparison of the 16S rRNA Gene Sequences of Species of the Genus Pseudomonas (sensu stricto and Estimation of the Natural Intrageneric Relationships.** *Systematic and Applied Microbiology* 1996, **19**(4):478-492.
119. Anzai Y, Kim H, Park JY, Wakabayashi H, Oyaizu H: **Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence.** *Int J Syst Evol Microbiol* 2000, **50 Pt 4**:1563-1589.
120. Vancanneyt M, Witt S, Abraham W-R, Kersters K, Fredrickson HL: **Fatty acid content in whole-cell hydrolysates and phospholipid and phospholipid fractions of Pseudomonads: a taxonomic evaluation.** *Systematic and Applied Microbiology* 1996, **19**(4):528-540.
121. Yamamoto S, Kasai H, Arnold DL, Jackson RW, Vivian A, Harayama S: **Phylogeny of the genus Pseudomonas: intrageneric structure reconstructed from the nucleotide sequences of gyrB and rpoD genes.** *Microbiology* 2000, **146 ( Pt 10)**:2385-2394.
122. Hilario E, Buckley TR, Young JM: **Improved resolution on the phylogenetic relationships among Pseudomonas by the combined analysis of atp D, car A, rec A and 16S rDNA.** *Antonie Van Leeuwenhoek* 2004, **86**(1):51-64.
123. Tayeb LA, Ageron E, Grimont F, Grimont PA: **Molecular phylogeny of the genus Pseudomonas based on rpoB sequences and application for the identification of isolates.** *Res Microbiol* 2005, **156**(5-6):763-773.
124. Matthijs S, Coorevits A, Gebrekidan TT, Tricot C, Wauven CV, Pirnay JP, De Vos P, Cornelis P: **Evaluation of oprI and oprL genes as molecular markers for the genus Pseudomonas and their use in studying the biodiversity of a small Belgian River.** *Res Microbiol* 2013, **164**(3):254-261.
125. Gomila M, Pena A, Mulet M, Lalucat J, Garcia-Valdes E: **Phylogenomics and systematics in Pseudomonas.** *Front Microbiol* 2015, **6**:214.
126. García-Valdés E, Lalucat J: **Pseudomonas: Molecular Phylogeny and Current Taxonomy.** In: *Pseudomonas: Molecular and Applied Biology.* Edited by Kahlon RS: Springer; 2016: 1-23.
127. Cladera AM, Bennasar A, Barcelo M, Lalucat J, Garcia-Valdes E: **Comparative genetic diversity of Pseudomonas stutzeri genomovars, clonal structure, and phylogeny of the species.** *J Bacteriol* 2004, **186**(16):5239-5248.
128. Mulet M, Gomila M, Scotta C, Sanchez D, Lalucat J, Garcia-Valdes E: **Concordance between whole-cell matrix-assisted laser-desorption/ionization time-of-flight mass**

- spectrometry and multilocus sequence analysis approaches in species discrimination within the genus *Pseudomonas***. *Syst Appl Microbiol* 2012, **35**(7):455-464.
129. Mulet M, Garcia-Valdes E, Lalucat J: **Phylogenetic affiliation of *Pseudomonas putida* biovar A and B strains**. *Res Microbiol* 2013, **164**(4):351-359.
130. Scotta C, Gomila M, Mulet M, Lalucat J, Garcia-Valdes E: **Whole-cell MALDI-TOF mass spectrometry and multilocus sequence analysis in the discrimination of *Pseudomonas stutzeri* populations: three novel genomovars**. *Microb Ecol* 2013, **66**(3):522-532.
131. Marcelletti S, Scortichini M: **Definition of Plant-Pathogenic *Pseudomonas* Genomespecies of the *Pseudomonas syringae* Complex Through Multiple Comparative Approaches**. *Phytopathology* 2014, **104**(12):1274-1282.
132. Garrido-Sanz D, Meier-Kolthoff JP, Goker M, Martin M, Rivilla R, Redondo-Nieto M: **Genomic and Genetic Diversity within the *Pseudomonas fluorescens* Complex**. *PLoS one* 2016, **11**(2):e0150183.
133. Tourkya B, Boubellouta T, Dufour E, Leriche F: **Fluorescence spectroscopy as a promising tool for a polyphasic approach to pseudomonad taxonomy**. *Curr Microbiol* 2009, **58**(1):39-46.
134. Meyer JM, Geoffroy VA, Baida N, Gardan L, Izard D, Lemanceau P, Achouak W, Palleroni NJ: **Siderophore typing, a powerful tool for the identification of fluorescent and nonfluorescent pseudomonads**. *Appl Environ Microbiol* 2002, **68**(6):2745-2753.
135. Meyer JM: **Siderotyping and bacterial taxonomy: a siderophore bank for a rapid identification at the species level of fluorescent and non-fluorescent *Pseudomonas***. In: *Soil Biology, Microbial Siderophores*. Edited by Varma A, Chincholkar SB, vol. 12: Springer, Berlin Heidelberg; 2007: 43-65.
136. Mulet M, Gomila M, Gruffaz C, Meyer JM, Palleroni NJ, Lalucat J, Garcia-Valdes E: **Phylogenetic analysis and siderotyping as useful tools in the taxonomy of *Pseudomonas stutzeri*: description of a novel genomovar**. *Int J Syst Evol Microbiol* 2008, **58**(Pt 10):2309-2315.
137. Mulet M, Gomila M, Ramirez A, Cardew S, Moore ER, Lalucat J, Garcia-Valdes E: **Uncommonly isolated clinical *Pseudomonas*: identification and phylogenetic assignation**. *Eur J Clin Microbiol Infect Dis* 2017, **36**(2):351-359.
138. Behrendt U, Ulrich A, Schumann P, Erler W, Burghardt J, Seyfarth W: **A taxonomic study of bacteria isolated from grasses: a proposed new species *Pseudomonas graminis* sp. nov.** *Int J Syst Bacteriol* 1999, **49 Pt 1**:297-308.
139. Peix A, Rivas R, Mateos PF, Martinez-Molina E, Rodriguez-Barrueco C, Velazquez E: ***Pseudomonas rhizosphaerae* sp. nov., a novel species that actively solubilizes phosphate in vitro**. *Int J Syst Evol Microbiol* 2003, **53**(Pt 6):2067-2072.
140. Zhang DC, Liu HC, Zhou YG, Schinner F, Margesin R: ***Pseudomonas bauzanensis* sp. nov., isolated from soil**. *Int J Syst Evol Microbiol* 2011, **61**(Pt 10):2333-2337.
141. Pascual J, Lucena T, Ruvira MA, Giordano A, Gambacorta A, Garay E, Arahall DR, Pujalte MJ, Macian MC: ***Pseudomonas litoralis* sp. nov., isolated from Mediterranean seawater**. *Int J Syst Evol Microbiol* 2012, **62**(Pt 2):438-444.
142. Reddy GS, Matsumoto GI, Schumann P, Stackebrandt E, Shivaji S: **Psychrophilic pseudomonads from Antarctica: *Pseudomonas antarctica* sp. nov., *Pseudomonas meridiana* sp. nov. and *Pseudomonas proteolytica* sp. nov.** *Int J Syst Evol Microbiol* 2004, **54**(Pt 3):713-719.
143. Campos VL, Valenzuela C, Yarza P, Kampfer P, Vidal R, Zaror C, Mondaca MA, Lopez-Lopez A, Rossello-Mora R: ***Pseudomonas arsenicoxydans* sp. nov., an arsenite-oxidizing strain isolated from the Atacama desert**. *Syst Appl Microbiol* 2010, **33**(4):193-197.
144. Liu YC, Young LS, Lin SY, Hameed A, Hsu YH, Lai WA, Shen FT, Young CC: ***Pseudomonas guguanensis* sp. nov., a gammaproteobacterium isolated from a hot spring**. *Int J Syst Evol Microbiol* 2013, **63**(Pt 12):4591-4598.
145. Wang MQ, Sun L: ***Pseudomonas oceani* sp. nov., isolated from deep seawater**. *Int J Syst Evol Microbiol* 2016, **66**(10):4250-4255.
146. Migula W: **System der Bakterien, Handbuch der Morphologie, Entwicklungsgeschichte und Systematik der Bakterien**. *Gustav Fisher, Jena, Germany* 1900, **2**.
147. Van Hall CJJ: **Bijdragen tot de kennis der Bakterieele Plantenziekten. (Inaugural Dissertation Amsterdam)**; 1902.

148. Paine SG: **Studies in bacteriosis. II. A brown blotch disease of cultivated mushrooms.** *Annals of Applied Biology* 1919, **5**:206-219.
149. Wakabayashi H, Egusa S: **Characteristics of a Pseudomonas sp. from an epizootic of pond-cultured eels (*Anguilla japonica*).** *Bulletin of the Japanese Society of Scientific Fisheries* 1972, **38**(6):577-587.
150. Young JM, Saddler GS, Takikawa Y, De Boer SH, Vauterin L, Gardan L, Gvozdyak RI, Stead DE: **Names of plant pathogenic bacteria 1864-1995.** *Review of Plant Pathology* 1996, **75**(9):721-763.
151. Stanier RY, Palleroni NJ, Doudoroff M: **The aerobic pseudomonads: a taxonomic study.** *J Gen Microbiol* 1966, **43**(2):159-271.
152. Singh PB, Saini HS, Kahlon RS: **Pseudomonas: The Versatile and Adaptive Metabolic Network.** In: *Pseudomonas: Molecular and Applied Biology*. Edited by Kahlon RS: Springer; 2016: 81-126.
153. Olivera ER, Carnicero D, Jodra R, Minambres B, Garcia B, Abraham GA, Gallardo A, Roman JS, Garcia JL, Naharro G *et al*: **Genetically engineered Pseudomonas: a factory of new bioplastics with broad applications.** *Environ Microbiol* 2001, **3**(10):612-618.
154. Haas D, Defago G: **Biological control of soil-borne pathogens by fluorescent pseudomonads.** *Nat Rev Microbiol* 2005, **3**(4):307-319.
155. Weller DM: **Pseudomonas biocontrol agents of soilborne pathogens: looking back over 30 years.** *Phytopathology* 2007, **97**(2):250-256.
156. Kahlon RS: **Pseudomonas for Industrial Biotechnology.** In: *Pseudomonas: Molecular and Applied Biology*. Edited by Kahlon RS: Springer; 2016: 281-342.
157. Kahlon RS: **Biodegradation and Bioremediation of Organic Pollutants by Pseudomonas.** In: *Pseudomonas: Molecular and Applied Biology*. Edited by Kahlon RS: Springer; 2016: 343-418.
158. Khan H, Parmar N, Kahlon RS: **Pseudomonas-Plant Interactions I: Plant Growth Promotion and Defense-Mediated Mechanisms.** In: *Pseudomonas: Molecular and Applied Biology*. Edited by Kahlon RS: Springer; 2016: 419-468.
159. Vilain S, Luo Y, Hildreth MB, Brozel VS: **Analysis of the life cycle of the soil saprophyte *Bacillus cereus* in liquid soil extract and in soil.** *Appl Environ Microbiol* 2006, **72**(7):4970-4977.
160. Ceuppens S, Boon N, Uyttendaele M: **Diversity of *Bacillus cereus* group strains is reflected in their broad range of pathogenicity and diverse ecological lifestyles.** *FEMS Microbiol Ecol* 2013, **84**(3):433-450.
161. Frankland GC, Frankland PF: **Studies on some new micro-organisms obtained from air.** *Philosophical Transactions of the Royal Society of London B* 1887, **178**:257-287.
162. Cohn F: **Untersuchungen über Bakterien.** *Beitrage zur Biologie der Pflanzen, 1875 1 (Heft 2)* 1872, **1**:127-224.
163. Berliner E: **Über die Schlaffsucht der Mehlmottenraupe (*Ephestia kühniella* Zell.) und ihren Erreger *Bacillus thuringiensis* n. sp.** *Journal of Applied Entomology* 1915, **2**(1):29-56.
164. Flügge C: **Die Mikroorganismen.** *FCW Vogel, Leipzig* 1886, **2**.
165. Nakamura LK: ***Bacillus pseudomycooides* sp. nov.** *Int J Syst Bacteriol* 1998, **48 Pt 3**:1031-1035.
166. Lechner S, Mayr R, Francis KP, Pruss BM, Kaplan T, Wiessner-Gunkel E, Stewart GS, Scherer S: ***Bacillus weihenstephanensis* sp. nov. is a new psychrotolerant species of the *Bacillus cereus* group.** *Int J Syst Bacteriol* 1998, **48 Pt 4**:1373-1382.
167. Guinebretiere MH, Auger S, Galleron N, Contzen M, De Sarrau B, De Buyser ML, Lamberet G, Fagerlund A, Granum PE, Lereclus D *et al*: ***Bacillus cytotoxicus* sp. nov. is a novel thermotolerant species of the *Bacillus cereus* Group occasionally associated with food poisoning.** *Int J Syst Evol Microbiol* 2013, **63**(Pt 1):31-40.
168. Jimenez G, Urdiain M, Cifuentes A, Lopez-Lopez A, Blanch AR, Tamames J, Kampfer P, Kolsto AB, Ramon D, Martinez JF *et al*: **Description of *Bacillus toyonensis* sp. nov., a novel species of the *Bacillus cereus* group, and pairwise genome comparisons of the species of the group by means of ANI calculations.** *Syst Appl Microbiol* 2013, **36**(6):383-391.

169. Miller RA, Beno SM, Kent DJ, Carroll LM, Martin NH, Boor KJ, Kovac J: **Bacillus wiedmannii sp. nov., a psychrotolerant and cytotoxic Bacillus cereus group species isolated from dairy foods and dairy environments.** *Int J Syst Evol Microbiol* 2016, **66**(11):4744-4753.
170. Jung MY, Paek WK, Park IS, Han JR, Sin Y, Paek J, Rhee MS, Kim H, Song HS, Chang YH: **Bacillus gaemokensis sp. nov., isolated from foreshore tidal flat sediment from the Yellow Sea.** *J Microbiol* 2010, **48**(6):867-871.
171. Jung MY, Kim JS, Paek WK, Lim J, Lee H, Kim PI, Ma JY, Kim W, Chang YH: **Bacillus manliponensis sp. nov., a new member of the Bacillus cereus group isolated from foreshore tidal flat sediment.** *J Microbiol* 2011, **49**(6):1027-1032.
172. Liu B, Liu GH, Hu GP, Sengonca C, Lin NQ, Tang JY, Tang WQ, Lin YZ: **Bacillus bingmayongensis sp. nov., isolated from the pit soil of Emperor Qin's Terra-cotta warriors in China.** *Antonie Van Leeuwenhoek* 2014, **105**(3):501-510.
173. Hong HA, Duc le H, Cutting SM: **The use of bacterial spore formers as probiotics.** *FEMS Microbiol Rev* 2005, **29**(4):813-835.
174. Lacey LA, Grzywacz D, Shapiro-Ilan DI, Frutos R, Brownbridge M, Goettel MS: **Insect pathogens as biological control agents: Back to the future.** *J Invertebr Pathol* 2015, **132**:1-41.
175. Koch R: **Die Aetiologie der Milzbrand-Krankheit, begründet auf die Entwicklungsgeschichte des Bacillus anthracis.** *Beitr Biol Pflanzen* 1876, **2**:277-310.
176. Mock M, Fouet A: **Anthrax.** *Annu Rev Microbiol* 2001, **55**:647-671.
177. Goel AK: **Anthrax: A disease of biowarfare and public health importance.** *World J Clin Cases* 2015, **3**(1):20-33.
178. Palma L, Munoz D, Berry C, Murillo J, Caballero P: **Bacillus thuringiensis toxins: an overview of their biocidal activity.** *Toxins (Basel)* 2014, **6**(12):3296-3325.
179. Jouzani GS, Valijanjan E, Sharafi R: **Bacillus thuringiensis: a successful insecticide with new environmental features and tidings.** *Appl Microbiol Biotechnol* 2017, **101**(7):2691-2711.
180. Kramer JM, Gilbert RJ: **Bacillus cereus and other Bacillus species.** *Foodborne bacterial pathogens* 1989, **19**:21-70.
181. Granum PE, Lund T: **Bacillus cereus and its food poisoning toxins.** *FEMS Microbiol Lett* 1997, **157**(2):223-228.
182. Ehling-Schulz M, Fricker M, Scherer S: **Bacillus cereus, the causative agent of an emetic type of food-borne illness.** *Mol Nutr Food Res* 2004, **48**(7):479-487.
183. Stenfors Arnesen LP, Fagerlund A, Granum PE: **From soil to gut: Bacillus cereus and its food poisoning toxins.** *FEMS Microbiol Rev* 2008, **32**(4):579-606.
184. Ehling-Schulz M, Knutsson R, Scherer S: **Bacillus cereus.** In: *Genomes of Foodborne and Waterborne Pathogens.* American Society of Microbiology; 2011: 147-164.
185. Tewari A, Abdullah S: **Bacillus cereus food poisoning: international and Indian perspective.** *J Food Sci Technol* 2015, **52**(5):2500-2511.
186. Mahler H, Pasi A, Kramer JM, Schulte P, Scoging AC, Bar W, Krahenbuhl S: **Fulminant liver failure in association with the emetic toxin of Bacillus cereus.** *N Engl J Med* 1997, **336**(16):1142-1148.
187. Dierick K, Van Coillie E, Swiecicka I, Meyfroidt G, Devlieger H, Meulemans A, Hoedemaekers G, Fourie L, Heyndrickx M, Mahillon J: **Fatal family outbreak of Bacillus cereus-associated food poisoning.** *J Clin Microbiol* 2005, **43**(8):4277-4279.
188. Posfay-Barbe KM, Schrenzel J, Frey J, Studer R, Korff C, Belli DC, Parvex P, Rimensberger PC, Schappi MG: **Food poisoning as a cause of acute liver failure.** *Pediatr Infect Dis J* 2008, **27**(9):846-847.
189. Naranjo M, Denayer S, Botteldoorn N, Delbrassinne L, Veys J, Waegenaere J, Sirtaine N, Driesen RB, Sipido KR, Mahillon J *et al*: **Sudden death of a young adult associated with Bacillus cereus food poisoning.** *J Clin Microbiol* 2011, **49**(12):4379-4381.
190. Tschiedel E, Rath PM, Steinmann J, Becker H, Dietrich R, Paul A, Felderhoff-Muser U, Dohna-Schwake C: **Lifesaving liver transplantation for multi-organ failure caused by Bacillus cereus food poisoning.** *Pediatr Transplant* 2015, **19**(1):E11-14.
191. EFSA, ECDC: **The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2011.** *EFSA Journal* 2013, **11**(4).

192. EFSA, ECDC: **The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2012.** *EFSA Journal* 2014, **12**(2).
193. EFSA, ECDC: **The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2013.** *EFSA Journal* 2015, **13**(1).
194. Helgason E, Caugant DA, Olsen I, Kolsto AB: **Genetic structure of population of *Bacillus cereus* and *B. thuringiensis* isolates associated with periodontitis and other human infections.** *J Clin Microbiol* 2000, **38**(4):1615-1622.
195. Horii T, Tamai K, Notake S, Yanagisawa H: ***Bacillus cereus* Bloodstream Infection in a Preterm Neonate Complicated by Late Meningitis.** *Case Rep Infect Dis* 2012, **2012**:358789.
196. Stevens MP, Elam K, Bearman G: **Meningitis due to *Bacillus cereus*: A case report and review of the literature.** *Can J Infect Dis Med Microbiol* 2012, **23**(1):e16-19.
197. Miyata J, Tasaka S, Miyazaki M, Yoshida S, Naoki K, Sayama K, Asano K, Fujiwara H, Ohkusu K, Hasegawa N *et al*: ***Bacillus cereus* necrotizing pneumonia in a patient with nephrotic syndrome.** *Intern Med* 2013, **52**(1):101-104.
198. Rishi E, Rishi P, Sengupta S, Jambulingam M, Madhavan HN, Gopal L, Therese KL: **Acute postoperative *Bacillus cereus* endophthalmitis mimicking toxic anterior segment syndrome.** *Ophthalmology* 2013, **120**(1):181-185.
199. Hong M, Wang Q, Tang Z, Wang Y, Gu Y, Lou Y, Zheng M: **Association of Genotyping of *Bacillus cereus* with Clinical Features of Post-Traumatic Endophthalmitis.** *PLoS one* 2016, **11**(2):e0147878.
200. Shimoyama Y, Umegaki O, Ooi Y, Agui T, Kadono N, Minami T: ***Bacillus cereus* pneumonia in an immunocompetent patient: a case report.** *JA Clin Rep* 2017, **3**(1):25.
201. Nakamura LK, Jackson MA: **Clarification of the taxonomy of *Bacillus mycoides*.** *International Journal of Systematic and Evolutionary Microbiology* 1995, **45**(1):46-49.
202. Ivy RA, Ranieri ML, Martin NH, den Bakker HC, Xavier BM, Wiedmann M, Boor KJ: **Identification and characterization of psychrotolerant sporeformers associated with fluid milk production and processing.** *Appl Environ Microbiol* 2012, **78**(6):1853-1864.
203. Markland SM, Farkas DF, Kniel KE, Hoover DG: **Pathogenic psychrotolerant sporeformers: an emerging challenge for low-temperature storage of minimally processed foods.** *Foodborne Pathog Dis* 2013, **10**(5):413-419.
204. Techer C, Baron F, Delbrassinne L, Belaid R, Brunet N, Gillard A, Gonnet F, Cochet MF, Grosset N, Gautier M *et al*: **Global overview of the risk linked to the *Bacillus cereus* group in the egg product industry: identification of food safety and food spoilage markers.** *J Appl Microbiol* 2014, **116**(5):1344-1358.
205. Doll EV, Scherer S, Wenning M: **Spoilage of Microfiltered and Pasteurized Extended Shelf Life Milk Is Mainly Induced by Psychrotolerant Spore-Forming Bacteria that often Originate from Recontamination.** *Front Microbiol* 2017, **8**:135.
206. Thorsen L, Hansen BM, Nielsen KF, Hendriksen NB, Phipps RK, Budde BB: **Characterization of emetic *Bacillus weihenstephanensis*, a new cereulide-producing bacterium.** *Appl Environ Microbiol* 2006, **72**(7):5118-5121.
207. Hoton FM, Fornelos N, N'Guessan E, Hu X, Swiecicka I, Dierick K, Jaaskelainen E, Salkinoja-Salonen M, Mahillon J: **Family portrait of *Bacillus cereus* and *Bacillus weihenstephanensis* cereulide-producing strains.** *Environ Microbiol Rep* 2009, **1**(3):177-183.
208. Lund T, De Buyser ML, Granum PE: **A new cytotoxin from *Bacillus cereus* that may cause necrotic enteritis.** *Mol Microbiol* 2000, **38**(2):254-261.
209. Beecher DJ, Olsen TW, Somers EB, Wong AC: **Evidence for contribution of tripartite hemolysin BL, phosphatidylcholine-preferring phospholipase C, and collagenase to virulence of *Bacillus cereus* endophthalmitis.** *Infect Immun* 2000, **68**(9):5269-5276.
210. Cutting SM: ***Bacillus* probiotics.** *Food Microbiol* 2011, **28**(2):214-220.
211. Selander RK, Caugant DA, Ochman H, Musser JM, Gilmour MN, Whittam TS: **Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics.** *Appl Environ Microbiol* 1986, **51**(5):873-884.
212. Somerville HJ, Jones ML: **DNA Competition Studies within the *Bacillus cereus* Group of Bacilli.** *Journal of General Microbiology* 1972, **73**(2):257-265.

213. Kaneko T, Nozaki R, Aizawa K: **Deoxyribonucleic Acid Relatedness between Bacillus anthracis, Bacillus cereus and Bacillus thuringiensis.** *Microbiology and immunology* 1978, **22**(10):639-641.
214. Seki T, Chung C-K, Mikami H, Oshima Y: **Deoxyribonucleic Acid Homology and Taxonomy of the Genus Bacillus.** *International Journal of Systematic and Evolutionary Microbiology* 1978, **28**(2):182-189.
215. Ash C, Farrow JA, Dorsch M, Stackebrandt E, Collins MD: **Comparative analysis of Bacillus anthracis, Bacillus cereus, and related species on the basis of reverse transcriptase sequencing of 16S rRNA.** *Int J Syst Bacteriol* 1991, **41**(3):343-346.
216. Carlson CR, Caugant DA, Kolsto AB: **Genotypic Diversity among Bacillus cereus and Bacillus thuringiensis Strains.** *Appl Environ Microbiol* 1994, **60**(6):1719-1725.
217. Helgason E, Okstad OA, Caugant DA, Johansen HA, Fouet A, Mock M, Hegna I, Kolsto AB: **Bacillus anthracis, Bacillus cereus, and Bacillus thuringiensis--one species on the basis of genetic evidence.** *Appl Environ Microbiol* 2000, **66**(6):2627-2630.
218. Okinaka RT, Keim P: **The Phylogeny of Bacillus cereus sensu lato.** *Microbiol Spectr* 2016, **4**(1).
219. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA *et al*: **Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms.** *Proc Natl Acad Sci U S A* 1998, **95**(6):3140-3145.
220. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M *et al*: **AFLP: a new technique for DNA fingerprinting.** *Nucleic Acids Res* 1995, **23**(21):4407-4414.
221. Helgason E, Tourasse NJ, Meisal R, Caugant DA, Kolsto AB: **Multilocus sequence typing scheme for bacteria of the Bacillus cereus group.** *Appl Environ Microbiol* 2004, **70**(1):191-201.
222. Hill KK, Ticknor LO, Okinaka RT, Asay M, Blair H, Bliss KA, Laker M, Pardington PE, Richardson AP, Tonks M *et al*: **Fluorescent amplified fragment length polymorphism analysis of Bacillus anthracis, Bacillus cereus, and Bacillus thuringiensis isolates.** *Appl Environ Microbiol* 2004, **70**(2):1068-1080.
223. Priest FG, Barker M, Baillie LW, Holmes EC, Maiden MC: **Population structure and evolution of the Bacillus cereus group.** *J Bacteriol* 2004, **186**(23):7959-7970.
224. Sorokin A, Candelon B, Guilloux K, Galleron N, Wackerow-Kouzova N, Ehrlich SD, Bourguet D, Sanchis V: **Multiple-locus sequence typing analysis of Bacillus cereus and Bacillus thuringiensis reveals separate clustering and a distinct population structure of psychrotrophic strains.** *Appl Environ Microbiol* 2006, **72**(2):1569-1578.
225. Hoffmaster AR, Novak RT, Marston CK, Gee JE, Hessel L, Pruckler JM, Wilkins PP: **Genetic diversity of clinical isolates of Bacillus cereus using multilocus sequence typing.** *BMC Microbiol* 2008, **8**:191.
226. Didelot X, Barker M, Falush D, Priest FG: **Evolution of pathogenicity in the Bacillus cereus group.** *Syst Appl Microbiol* 2009, **32**(2):81-90.
227. Harrell LJ, Andersen GL, Wilson KH: **Genetic variability of Bacillus anthracis and related species.** *J Clin Microbiol* 1995, **33**(7):1847-1850.
228. Keim P, Kalif A, Schupp J, Hill K, Travis SE, Richmond K, Adair DM, Hugh-Jones M, Kuske CR, Jackson P: **Molecular evolution and diversity in Bacillus anthracis as detected by amplified fragment length polymorphism markers.** *J Bacteriol* 1997, **179**(3):818-824.
229. Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, Okinaka R, Jackson PJ, Hugh-Jones ME: **Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within Bacillus anthracis.** *J Bacteriol* 2000, **182**(10):2928-2936.
230. Schnepf E, Crickmore N, Van Rie J, Lereclus D, Baum J, Feitelson J, Zeigler DR, Dean DH: **Bacillus thuringiensis and its pesticidal crystal proteins.** *Microbiol Mol Biol Rev* 1998, **62**(3):775-806.
231. de Maagd RA, Bravo A, Berry C, Crickmore N, Schnepf HE: **Structure, diversity, and evolution of protein toxins from spore-forming entomopathogenic bacteria.** *Annu Rev Genet* 2003, **37**:409-433.
232. Damgaard PH: **Diarrhoeal enterotoxin production by strains of Bacillus thuringiensis isolated from commercial Bacillus thuringiensis-based insecticides.** *FEMS Immunol Med Microbiol* 1995, **12**(3-4):245-250.

233. Gaviria Rivera AM, Granum PE, Priest FG: **Common occurrence of enterotoxin genes and enterotoxicity in *Bacillus thuringiensis***. *FEMS microbiology letters* 2000, **190**(1):151-155.
234. Swiecicka I, Van der Auwera GA, Mahillon J: **Hemolytic and nonhemolytic enterotoxin genes are broadly distributed among *Bacillus thuringiensis* isolated from wild mammals**. *Microb Ecol* 2006, **52**(3):544-551.
235. Jackson SG, Goodbrand RB, Ahmed R, Kasatiya S: ***Bacillus cereus* and *Bacillus thuringiensis* isolated in a gastroenteritis outbreak investigation**. *Lett Appl Microbiol* 1995, **21**(2):103-105.
236. Damgaard PH, Granum PE, Bresciani J, Torregrossa MV, Eilenberg J, Valentino L: **Characterization of *Bacillus thuringiensis* isolated from infections in burn wounds**. *FEMS Immunol Med Microbiol* 1997, **18**(1):47-53.
237. McIntyre L, Bernard K, Beniac D, Isaac-Renton JL, Naseby DC: **Identification of *Bacillus cereus* group species associated with food poisoning outbreaks in British Columbia, Canada**. *Appl Environ Microbiol* 2008, **74**(23):7451-7453.
238. Kuroki R, Kawakami K, Qin L, Kaji C, Watanabe K, Kimura Y, Ishiguro C, Tanimura S, Tsuchiya Y, Hamaguchi I *et al*: **Nosocomial bacteremia caused by biofilm-forming *Bacillus cereus* and *Bacillus thuringiensis***. *Intern Med* 2009, **48**(10):791-796.
239. Koehler TM: ***Bacillus anthracis* physiology and genetics**. *Mol Aspects Med* 2009, **30**(6):386-396.
240. Kolsto AB, Tourasse NJ, Okstad OA: **What sets *Bacillus anthracis* apart from other *Bacillus* species?** *Annu Rev Microbiol* 2009, **63**:451-476.
241. Hoffmaster AR, Ravel J, Rasko DA, Chapman GD, Chute MD, Marston CK, De BK, Sacchi CT, Fitzgerald C, Mayer LW *et al*: **Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax**. *Proc Natl Acad Sci U S A* 2004, **101**(22):8449-8454.
242. Hoffmaster AR, Hill KK, Gee JE, Marston CK, De BK, Popovic T, Sue D, Wilkins PP, Avashia SB, Drumgoole R *et al*: **Characterization of *Bacillus cereus* isolates associated with fatal pneumonias: strains are closely related to *Bacillus anthracis* and harbor *B. anthracis* virulence genes**. *J Clin Microbiol* 2006, **44**(9):3352-3360.
243. Klee SR, Ozel M, Appel B, Boesch C, Ellerbrok H, Jacob D, Holland G, Leendertz FH, Pauli G, Grunow R *et al*: **Characterization of *Bacillus anthracis*-like bacteria isolated from wild great apes from Cote d'Ivoire and Cameroon**. *J Bacteriol* 2006, **188**(15):5333-5344.
244. Marston CK, Ibrahim H, Lee P, Churchwell G, Gumke M, Stanek D, Gee JE, Boyer AE, Gallegos-Candela M, Barr JR *et al*: **Anthrax Toxin-Expressing *Bacillus cereus* Isolated from an Anthrax-Like Eschar**. *PLoS one* 2016, **11**(6):e0156987.
245. Agaisse H, Gominet M, Okstad OA, Kolsto AB, Lereclus D: **PicR is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis***. *Mol Microbiol* 1999, **32**(5):1043-1053.
246. Mignot T, Mock M, Robichon D, Landier A, Lereclus D, Fouet A: **The incompatibility between the PicR- and AtxA-controlled regulons may have selected a nonsense mutation in *Bacillus anthracis***. *Mol Microbiol* 2001, **42**(5):1189-1198.
247. Gohar M, Faegri K, Perchat S, Ravnum S, Okstad OA, Gominet M, Kolsto AB, Lereclus D: **The PicR virulence regulon of *Bacillus cereus***. *PLoS one* 2008, **3**(7):e2793.
248. Ehling-Schulz M, Svensson B, Guinebretiere MH, Lindback T, Andersson M, Schulz A, Fricker M, Christiansson A, Granum PE, Martlbauer E *et al*: **Emetic toxin formation of *Bacillus cereus* is restricted to a single evolutionary lineage of closely related strains**. *Microbiology* 2005, **151**(Pt 1):183-197.
249. Vassileva M, Torii K, Oshimoto M, Okamoto A, Agata N, Yamada K, Hasegawa T, Ohta M: **A new phylogenetic cluster of cereulide-producing *Bacillus cereus* strains**. *J Clin Microbiol* 2007, **45**(4):1274-1277.
250. Ehling-Schulz M, Frenzel E, Gohar M: **Food-bacteria interplay: pathometabolism of emetic *Bacillus cereus***. *Front Microbiol* 2015, **6**:704.
251. Mei X, Xu K, Yang L, Yuan Z, Mahillon J, Hu X: **The genetic diversity of cereulide biosynthesis gene cluster indicates a composite transposon Tnces in emetic *Bacillus weihenstephanensis***. *BMC Microbiol* 2014, **14**:149.
252. Guinebretiere MH, Thompson FL, Sorokin A, Normand P, Dawyndt P, Ehling-Schulz M, Svensson B, Sanchis V, Nguyen-The C, Heyndrickx M *et al*: **Ecological diversification in the *Bacillus cereus* Group**. *Environ Microbiol* 2008, **10**(4):851-865.

253. Bazinet AL: **Pan-genome and phylogeny of *Bacillus cereus sensu lato***. *BMC Evol Biol* 2017, **17**(1):176.
254. Guinebretiere MH, Velge P, Couvert O, Carlin F, Debuyser ML, Nguyen-The C: **Ability of *Bacillus cereus* group strains to cause food poisoning varies according to phylogenetic affiliation (groups I to VII) rather than species affiliation**. *J Clin Microbiol* 2010, **48**(9):3388-3391.
255. Miller RA, Jian J, Beno SM, Wiedmann M, Kovac J: **Intraclade Variability in Toxin Production and Cytotoxicity of *Bacillus cereus* Group Type Strains and Dairy-Associated Isolates**. *Appl Environ Microbiol* 2018, **84**(6).
256. Ceuppens S, Rajkovic A, Heyndrickx M, Tsilia V, Van De Wiele T, Boon N, Uyttendaele M: **Regulation of toxin production by *Bacillus cereus* and its food safety implications**. *Crit Rev Microbiol* 2011, **37**(3):188-213.
257. Leppla SH: **The anthrax toxin complex**. *Source book of bacterial protein toxins* 1991, **14**:277-302.
258. Robertson DL, Tippetts MT, Leppla SH: **Nucleotide sequence of the *Bacillus anthracis* edema factor gene (*cya*): a calmodulin-dependent adenylate cyclase**. *Gene* 1988, **73**(2):363-371.
259. Welkos SL, Lowe JR, Eden-McCutchan F, Vodkin M, Leppla SH, Schmidt JJ: **Sequence and analysis of the DNA encoding protective antigen of *Bacillus anthracis***. *Gene* 1988, **69**(2):287-300.
260. Bragg TS, Robertson DL: **Nucleotide sequence and analysis of the lethal factor gene (*lef*) from *Bacillus anthracis***. *Gene* 1989, **81**(1):45-54.
261. Collier RJ, Young JA: **Anthrax toxin**. *Annu Rev Cell Dev Biol* 2003, **19**:45-70.
262. Young JA, Collier RJ: **Anthrax toxin: receptor binding, internalization, pore formation, and translocation**. *Annu Rev Biochem* 2007, **76**:243-265.
263. Duesbery NS, Webb CP, Leppla SH, Gordon VM, Klimpel KR, Copeland TD, Ahn NG, Oskarsson MK, Fukasawa K, Paull KD *et al*: **Proteolytic inactivation of MAP-kinase-kinase by anthrax lethal factor**. *Science* 1998, **280**(5364):734-737.
264. Vitale G, Pellizzari R, Recchi C, Napolitani G, Mock M, Montecucco C: **Anthrax lethal factor cleaves the N-terminus of MAPKKs and induces tyrosine/threonine phosphorylation of MAPKs in cultured macrophages**. *Biochem Biophys Res Commun* 1998, **248**(3):706-711.
265. Vitale G, Bernardi L, Napolitani G, Mock M, Montecucco C: **Susceptibility of mitogen-activated protein kinase kinase family members to proteolysis by anthrax lethal factor**. *Biochem J* 2000, **352 Pt 3**:739-745.
266. Leppla SH: **Anthrax toxin edema factor: a bacterial adenylate cyclase that increases cyclic AMP concentrations of eukaryotic cells**. *Proc Natl Acad Sci U S A* 1982, **79**(10):3162-3166.
267. Leppla SH: ***Bacillus anthracis* calmodulin-dependent adenylate cyclase: chemical and enzymatic properties and interactions with eucaryotic cells**. *Adv Cyclic Nucleotide Protein Phosphorylation Res* 1984, **17**:189-198.
268. Shen Y, Lee YS, Soelaiman S, Bergson P, Lu D, Chen A, Beckingham K, Grabarek Z, Mrksich M, Tang WJ: **Physiological calcium concentrations regulate calmodulin binding and catalysis of adenylate cyclase exotoxins**. *EMBO J* 2002, **21**(24):6721-6732.
269. Liu S, Moayeri M, Leppla SH: **Anthrax lethal and edema toxins in anthrax pathogenesis**. *Trends Microbiol* 2014, **22**(6):317-325.
270. Friebe S, van der Goot FG, Burgi J: **The Ins and Outs of Anthrax Toxin**. *Toxins (Basel)* 2016, **8**(3).
271. Raymond B, Johnston PR, Nielsen-LeRoux C, Lereclus D, Crickmore N: ***Bacillus thuringiensis*: an impotent pathogen?** *Trends Microbiol* 2010, **18**(5):189-194.
272. Crickmore N, Zeigler DR, Feitelson J, Schnepf E, Van Rie J, Lereclus D, Baum J, Dean DH: **Revision of the nomenclature for the *Bacillus thuringiensis* pesticidal crystal proteins**. *Microbiol Mol Biol Rev* 1998, **62**(3):807-813.
273. Donovan WP, Engleman JT, Donovan JC, Baum JA, Bunkers GJ, Chi DJ, Clinton WP, English L, Heck GR, Ilagan OM *et al*: **Discovery and characterization of Sip1A: A novel secreted protein from *Bacillus thuringiensis* with activity against coleopteran larvae**. *Appl Microbiol Biotechnol* 2006, **72**(4):713-719.

274. Agata N, Mori M, Ohta M, Suwan S, Ohtani I, Isobe M: **A novel dodecadepsipeptide, cereulide, isolated from *Bacillus cereus* causes vacuole formation in HEP-2 cells.** *FEMS Microbiol Lett* 1994, **121**(1):31-34.
275. Melling J, Capel BJ, Turnbull PC, Gilbert RJ: **Identification of a novel enterotoxigenic activity associated with *Bacillus cereus*.** *J Clin Pathol* 1976, **29**(10):938-940.
276. Melling J, Capel BJ: **Characteristics of *Bacillus cereus* emetic toxin.** *FEMS Microbiology Letters* 1978, **4**(3):133-135.
277. Shinagawa K, Konuma H, Sekita H, Sugii S: **Emesis of rhesus monkeys induced by intragastric administration with the HEP-2 vacuolation factor (cereulide) produced by *Bacillus cereus*.** *FEMS Microbiol Lett* 1995, **130**(1):87-90.
278. Rajkovic A, Uyttendaele M, Vermeulen A, Andjelkovic M, Fitz-James I, in 't Veld P, Denon Q, Verhe R, Debevere J: **Heat resistance of *Bacillus cereus* emetic toxin, cereulide.** *Lett Appl Microbiol* 2008, **46**(5):536-541.
279. Agata N, Ohta M, Yokoyama K: **Production of *Bacillus cereus* emetic toxin (cereulide) in various foods.** *Int J Food Microbiol* 2002, **73**(1):23-27.
280. Agata N, Ohta M, Mori M, Isobe M: **A novel dodecadepsipeptide, cereulide, is an emetic toxin of *Bacillus cereus*.** *FEMS Microbiol Lett* 1995, **129**(1):17-20.
281. Ladeuze S, Lentz N, Delbrassinne L, Hu X, Mahillon J: **Antifungal activity displayed by cereulide, the emetic toxin produced by *Bacillus cereus*.** *Appl Environ Microbiol* 2011, **77**(7):2555-2558.
282. Tempelaars MH, Rodrigues S, Abee T: **Comparative analysis of antimicrobial activities of valinomycin and cereulide, the *Bacillus cereus* emetic toxin.** *Appl Environ Microbiol* 2011, **77**(8):2755-2762.
283. Mikkola R, Saris NE, Grigoriev PA, Andersson MA, Salkinoja-Salonen MS: **Ionophoretic properties and mitochondrial effects of cereulide: the emetic toxin of *B. cereus*.** *Eur J Biochem* 1999, **263**(1):112-117.
284. Ekman JV, Kruglov A, Andersson MA, Mikkola R, Raulio M, Salkinoja-Salonen M: **Cereulide produced by *Bacillus cereus* increases the fitness of the producer organism in low-potassium environments.** *Microbiology* 2012, **158**(Pt 4):1106-1116.
285. Horwood PF, Burgess GW, Oakey HJ: **Evidence for non-ribosomal peptide synthetase production of cereulide (the emetic toxin) in *Bacillus cereus*.** *FEMS Microbiol Lett* 2004, **236**(2):319-324.
286. Ehling-Schulz M, Vukov N, Schulz A, Shaheen R, Andersson M, Martlbauer E, Scherer S: **Identification and partial characterization of the nonribosomal peptide synthetase gene responsible for cereulide production in emetic *Bacillus cereus*.** *Appl Environ Microbiol* 2005, **71**(1):105-113.
287. Ehling-Schulz M, Fricker M, Grallert H, Rieck P, Wagner M, Scherer S: **Cereulide synthetase gene cluster from emetic *Bacillus cereus*: structure and location on a mega virulence plasmid related to *Bacillus anthracis* toxin plasmid pXO1.** *BMC Microbiol* 2006, **6**:20.
288. Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M, Marahiel MA, Reid R, Khosla C, Walsh CT: **A new enzyme superfamily - the phosphopantetheinyl transferases.** *Chem Biol* 1996, **3**(11):923-936.
289. Schwarzer D, Mootz HD, Linne U, Marahiel MA: **Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases.** *Proc Natl Acad Sci U S A* 2002, **99**(22):14083-14088.
290. Frenzel E, Doll V, Pauthner M, Lucking G, Scherer S, Ehling-Schulz M: **CodY orchestrates the expression of virulence determinants in emetic *Bacillus cereus* by impacting key regulatory circuits.** *Mol Microbiol* 2012, **85**(1):67-88.
291. Lucking G, Dommel MK, Scherer S, Fouet A, Ehling-Schulz M: **Cereulide synthesis in emetic *Bacillus cereus* is controlled by the transition state regulator AbrB, but not by the virulence regulator PlcR.** *Microbiology* 2009, **155**(Pt 3):922-931.
292. Senesi S, Ghelardi E: **Production, secretion and biological activity of *Bacillus cereus* enterotoxins.** *Toxins (Basel)* 2010, **2**(7):1690-1703.
293. Lund T, Granum PE: **Characterisation of a non-haemolytic enterotoxin complex from *Bacillus cereus* isolated after a foodborne outbreak.** *FEMS Microbiol Lett* 1996, **141**(2-3):151-156.
294. Beecher DJ, Macmillan JD: **Characterization of the components of hemolysin BL from *Bacillus cereus*.** *Infect Immun* 1991, **59**(5):1778-1784.

295. Heinrichs JH, Beecher DJ, MacMillan JD, Zilinskas BA: **Molecular cloning and characterization of the hblA gene encoding the B component of hemolysin BL from *Bacillus cereus*.** *J Bacteriol* 1993, **175**(21):6760-6766.
296. Ryan PA, Macmillan JD, Zilinskas BA: **Molecular cloning and characterization of the genes encoding the L1 and L2 components of hemolysin BL from *Bacillus cereus*.** *J Bacteriol* 1997, **179**(8):2551-2556.
297. Granum PE, O'Sullivan K, Lund T: **The sequence of the non-haemolytic enterotoxin operon from *Bacillus cereus*.** *FEMS Microbiol Lett* 1999, **177**(2):225-229.
298. Lindback T, Hardy SP, Dietrich R, Sodring M, Didier A, Moravek M, Fagerlund A, Bock S, Nielsen C, Casteel M *et al*: **Cytotoxicity of the *Bacillus cereus* Nhe enterotoxin requires specific binding order of its three exoprotein components.** *Infect Immun* 2010, **78**(9):3813-3821.
299. Lindback T, Fagerlund A, Rodland MS, Granum PE: **Characterization of the *Bacillus cereus* Nhe enterotoxin.** *Microbiology* 2004, **150**(Pt 12):3959-3967.
300. Moravek M, Dietrich R, Buerk C, Broussolle V, Guinebretiere MH, Granum PE, Nguyen-The C, Martlbauer E: **Determination of the toxic potential of *Bacillus cereus* isolates by quantitative enterotoxin analyses.** *FEMS Microbiol Lett* 2006, **257**(2):293-298.
301. Ngamwongsatit P, Buasri W, Pianariyanon P, Pulsrikarn C, Ohba M, Assavanig A, Panbangred W: **Broad distribution of enterotoxin genes (hblCDA, nheABC, cytK, and entFM) among *Bacillus thuringiensis* and *Bacillus cereus* as shown by novel primers.** *Int J Food Microbiol* 2008, **121**(3):352-356.
302. Lapidus A, Goltsman E, Auger S, Galleron N, Segurens B, Dossat C, Land ML, Broussolle V, Brillard J, Guinebretiere MH *et al*: **Extending the *Bacillus cereus* group genomics to putative food-borne pathogens of different toxicity.** *Chem Biol Interact* 2008, **171**(2):236-249.
303. Beecher DJ, MacMillan JD: **A novel bicomponent hemolysin from *Bacillus cereus*.** *Infect Immun* 1990, **58**(7):2220-2227.
304. Beecher DJ, Schoeni JL, Wong AC: **Enterotoxic activity of hemolysin BL from *Bacillus cereus*.** *Infect Immun* 1995, **63**(11):4423-4428.
305. Okstad OA, Gominet M, Purnelle B, Rose M, Lereclus D, Kolsto AB: **Sequence analysis of three *Bacillus cereus* loci carrying PlcR-regulated genes encoding degradative enzymes and enterotoxin.** *Microbiology* 1999, **145** ( Pt 11):3129-3138.
306. Beecher DJ, Wong AC: **Tripartite haemolysin BL: isolation and characterization of two distinct homologous sets of components from a single *Bacillus cereus* isolate.** *Microbiology* 2000, **146** ( Pt 6):1371-1380.
307. Duport C, Zigha A, Rosenfeld E, Schmitt P: **Control of enterotoxin gene expression in *Bacillus cereus* F4430/73 involves the redox-sensitive ResDE signal transduction system.** *J Bacteriol* 2006, **188**(18):6640-6651.
308. Esbelin J, Armengaud J, Zigha A, Duport C: **ResDE-dependent regulation of enterotoxin gene expression in *Bacillus cereus*: evidence for multiple modes of binding for ResD and interaction with Fnr.** *J Bacteriol* 2009, **191**(13):4419-4426.
309. Zigha A, Rosenfeld E, Schmitt P, Duport C: **The redox regulator Fnr is required for fermentative growth and enterotoxin synthesis in *Bacillus cereus* F4430/73.** *J Bacteriol* 2007, **189**(7):2813-2824.
310. Messaoudi K, Clavel T, Schmitt P, Duport C: **Fnr mediates carbohydrate-dependent regulation of catabolic and enterotoxin genes in *Bacillus cereus* F4430/73.** *Res Microbiol* 2010, **161**(1):30-39.
311. van der Voort M, Kuipers OP, Buist G, de Vos WM, Abee T: **Assessment of CcpA-mediated catabolite control of gene expression in *Bacillus cereus* ATCC 14579.** *BMC Microbiol* 2008, **8**:62.
312. Fagerlund A, Ween O, Lund T, Hardy SP, Granum PE: **Genetic and functional analysis of the cytK family of genes in *Bacillus cereus*.** *Microbiology* 2004, **150**(Pt 8):2689-2697.
313. Fagerlund A, Brillard J, Furst R, Guinebretiere MH, Granum PE: **Toxin production in a rare and genetically remote cluster of strains of the *Bacillus cereus* group.** *BMC Microbiol* 2007, **7**:43.
314. Ramarao N, Sanchis V: **The pore-forming haemolysins of *Bacillus cereus*: a review.** *Toxins (Basel)* 2013, **5**(6):1119-1139.

315. Guinebretiere MH, Fagerlund A, Granum PE, Nguyen-The C: **Rapid discrimination of *cytK-1* and *cytK-2* genes in *Bacillus cereus* strains by a novel duplex PCR system.** *FEMS Microbiol Lett* 2006, **259**(1):74-80.
316. Fagerlund A, Lindback T, Granum PE: ***Bacillus cereus* cytotoxins Hbl, Nhe and CytK are secreted via the Sec translocation pathway.** *BMC Microbiol* 2010, **10**:304.
317. Sinev MA, Budarina Zh I, Gavrilenko IV, Tomashevskii A, Kuz'min NP: **[Evidence of the existence of hemolysin II from *Bacillus cereus*: cloning the genetic determinant of hemolysin II].** *Mol Biol (Mosk)* 1993, **27**(6):1218-1229.
318. Baida G, Budarina ZI, Kuzmin NP, Solonin AS: **Complete nucleotide sequence and molecular characterization of hemolysin II gene from *Bacillus cereus*.** *FEMS Microbiol Lett* 1999, **180**(1):7-14.
319. Budarina ZI, Nikitin DV, Zenkin N, Zakharova M, Semenova E, Shlyapnikov MG, Rodikova EA, Masyukova S, Ogarkov O, Baida GE *et al*: **A new *Bacillus cereus* DNA-binding protein, HlyIIIR, negatively regulates expression of *B. cereus* haemolysin II.** *Microbiology* 2004, **150**(Pt 11):3691-3701.
320. Gohar M, Gilois N, Graveline R, Garreau C, Sanchis V, Lereclus D: **A comparative study of *Bacillus cereus*, *Bacillus thuringiensis* and *Bacillus anthracis* extracellular proteomes.** *Proteomics* 2005, **5**(14):3696-3711.
321. Harvie DR, Vilchez S, Steggle JR, Ellar DJ: ***Bacillus cereus* Fur regulates iron metabolism and is required for full virulence.** *Microbiology* 2005, **151**(Pt 2):569-577.
322. Sineva E, Shadrin A, Rodikova EA, Andreeva-Kovalevskaya ZI, Protsenko AS, Mayorov SG, Galaktionova DY, Magelky E, Solonin AS: **Iron regulates expression of *Bacillus cereus* hemolysin II via global regulator Fur.** *J Bacteriol* 2012, **194**(13):3327-3335.
323. Clair G, Roussi S, Armengaud J, Duport C: **Expanding the known repertoire of virulence factors produced by *Bacillus cereus* through early secretome profiling in three redox conditions.** *Mol Cell Proteomics* 2010, **9**(7):1486-1498.
324. Tran SL, Ramarao N: ***Bacillus cereus* immune escape: a journey within macrophages.** *FEMS Microbiol Lett* 2013, **347**(1):1-6.
325. Tran SL, Guillemet E, Ngo-Camus M, Clybouw C, Puhar A, Moris A, Gohar M, Lereclus D, Ramarao N: **Haemolysin II is a *Bacillus cereus* virulence factor that induces apoptosis of macrophages.** *Cell Microbiol* 2011, **13**(1):92-108.
326. Kreft J, Berger H, Hartlein M, Muller B, Weidinger G, Goebel W: **Cloning and expression in *Escherichia coli* and *Bacillus subtilis* of the hemolysin (cereolysin) determinant from *Bacillus cereus*.** *J Bacteriol* 1983, **155**(2):681-689.
327. Shannon JG, Ross CL, Koehler TM, Rest RF: **Characterization of anthrolysin O, the *Bacillus anthracis* cholesterol-dependent cytotoxin.** *Infect Immun* 2003, **71**(6):3183-3189.
328. Baida GE, Kuzmin NP: **Cloning and primary structure of a new hemolysin gene from *Bacillus cereus*.** *Biochim Biophys Acta* 1995, **1264**(2):151-154.
329. Dalhammar G, Steiner H: **Characterization of inhibitor A, a protease from *Bacillus thuringiensis* which degrades attacins and cecropins, two classes of antibacterial proteins in insects.** *Eur J Biochem* 1984, **139**(2):247-252.
330. Fedhila S, Nel P, Lereclus D: **The *InhA2* metalloprotease of *Bacillus thuringiensis* strain 407 is required for pathogenicity in insects infected via the oral route.** *J Bacteriol* 2002, **184**(12):3296-3304.
331. Guillemet E, Cadot C, Tran SL, Guinebretiere MH, Lereclus D, Ramarao N: **The *InhA* metalloproteases of *Bacillus cereus* contribute concomitantly to virulence.** *J Bacteriol* 2010, **192**(1):286-294.
332. Ramarao N, Lereclus D: **The *InhA1* metalloprotease allows spores of the *B. cereus* group to escape macrophages.** *Cell Microbiol* 2005, **7**(9):1357-1364.
333. Chung MC, Popova TG, Millis BA, Mukherjee DV, Zhou W, Liotta LA, Petricoin EF, Chandhoke V, Bailey C, Popov SG: **Secreted neutral metalloproteases of *Bacillus anthracis* as candidate pathogenic factors.** *J Biol Chem* 2006, **281**(42):31408-31418.
334. Asano SI, Nukumizu Y, Bando H, Iizuka T, Yamamoto T: **Cloning of novel enterotoxin genes from *Bacillus cereus* and *Bacillus thuringiensis*.** *Appl Environ Microbiol* 1997, **63**(3):1054-1057.
335. Boonchai N, Asano SI, Bando H, Wiwat C: **Study on cytotoxicity and nucleotide sequences of enterotoxin FM of *Bacillus cereus* isolated from various food sources.** *J Med Assoc Thai* 2008, **91**(9):1425-1432.

336. Tran SL, Guillemet E, Gohar M, Lereclus D, Ramarao N: **CwpFM (EntFM) is a Bacillus cereus potential cell wall peptidase implicated in adhesion, biofilm formation, and virulence.** *J Bacteriol* 2010, **192**(10):2638-2642.
337. Ikezawa H, Yamanegi M, Taguchi R, Miyashita T, Ohyabu T: **Studies on phosphatidylinositol phosphodiesterase (phospholipase C type) of Bacillus cereus. I. purification, properties and phosphatase-releasing activity.** *Biochim Biophys Acta* 1976, **450**(2):154-164.
338. Ikezawa H, Mori M, Ohyabu T, Taguchi R: **Studies on sphingomyelinase of Bacillus cereus. I. Purification and properties.** *Biochim Biophys Acta* 1978, **528**(2):247-256.
339. Gilmore MS, Cruz-Rodz AL, Leimeister-Wachter M, Kreft J, Goebel W: **A Bacillus cereus cytolytic determinant, cereolysin AB, which comprises the phospholipase C and sphingomyelinase genes: nucleotide sequence and genetic linkage.** *J Bacteriol* 1989, **171**(2):744-753.
340. Kuppe A, Evans LM, McMillen DA, Griffith OH: **Phosphatidylinositol-specific phospholipase C of Bacillus cereus: cloning, sequencing, and relationship to other phospholipases.** *J Bacteriol* 1989, **171**(11):6077-6083.
341. Doll VM, Ehling-Schulz M, Vogelmann R: **Concerted action of sphingomyelinase and non-hemolytic enterotoxin in pathogenic Bacillus cereus.** *PLoS one* 2013, **8**(4):e61404.
342. Huptas C, Scherer S, Wenning M: **Optimized Illumina PCR-free library preparation for bacterial whole genome sequencing and analysis of factors influencing de novo assembly.** *BMC Res Notes* 2016, **9**:269.
343. von Neubeck M, Huptas C, Gluck C, Krewinkel M, Stoeckel M, Stressler T, Fischer L, Hinrichs J, Scherer S, Wenning M: **Pseudomonas helleri sp. nov. and Pseudomonas weihenstephanensis sp. nov., isolated from raw cow's milk.** *Int J Syst Evol Microbiol* 2016, **66**(3):1163-1173.
344. von Neubeck M, Huptas C, Gluck C, Krewinkel M, Stoeckel M, Stressler T, Fischer L, Hinrichs J, Scherer S, Wenning M: **Pseudomonas lactis sp. nov. and Pseudomonas paralactis sp. nov., isolated from bovine raw milk.** *Int J Syst Evol Microbiol* 2017, **67**(6):1656-1664.
345. Kyrpides NC, Hugenholtz P, Eisen JA, Woyke T, Goker M, Parker CT, Amann R, Beck BJ, Chain PS, Chun J *et al*: **Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains.** *PLoS Biol* 2014, **12**(8):e1001920.
346. Kyrpides NC, Woyke T, Eisen JA, Garrity G, Lilburn TG, Beck BJ, Whitman WB, Hugenholtz P, Klenk HP: **Genomic Encyclopedia of Type Strains, Phase I: The one thousand microbial genomes (KMG-I) project.** *Stand Genomic Sci* 2014, **9**(3):1278-1284.
347. Whitman WB, Woyke T, Klenk HP, Zhou Y, Lilburn TG, Beck BJ, De Vos P, Vandamme P, Eisen JA, Garrity G *et al*: **Genomic Encyclopedia of Bacterial and Archaeal Type Strains, Phase III: the genomes of soil and plant-associated and newly described type strains.** *Stand Genomic Sci* 2015, **10**:26.
348. Spencer DH, Kas A, Smith EE, Raymond CK, Sims EH, Hastings M, Burns JL, Kaul R, Olson MV: **Whole-genome sequence variation among multiple isolates of Pseudomonas aeruginosa.** *J Bacteriol* 2003, **185**(4):1316-1325.
349. Wolfgang MC, Kulasekara BR, Liang X, Boyd D, Wu K, Yang Q, Miyada CG, Lory S: **Conservation of genome content and virulence determinants among clinical and environmental isolates of Pseudomonas aeruginosa.** *Proc Natl Acad Sci U S A* 2003, **100**(14):8484-8489.
350. Mathee K, Narasimhan G, Valdes C, Qiu X, Mawish JM, Koehrsen M, Rokas A, Yandava CN, Engels R, Zeng E *et al*: **Dynamics of Pseudomonas aeruginosa genome evolution.** *Proc Natl Acad Sci U S A* 2008, **105**(8):3100-3105.
351. Ozer EA, Allen JP, Hauser AR: **Characterization of the core and accessory genomes of Pseudomonas aeruginosa using bioinformatic tools Spine and AGent.** *BMC Genomics* 2014, **15**:737.
352. Keim P, Gruendike JM, Klevytska AM, Schupp JM, Challacombe J, Okinaka R: **The genome and variation of Bacillus anthracis.** *Mol Aspects Med* 2009, **30**(6):397-405.
353. Murray MG, Thompson WF: **Rapid isolation of high molecular weight plant DNA.** *Nucleic Acids Res* 1980, **8**(19):4321-4325.
354. **UCI Genomics High Throughput Facility** [<http://ghtf.biochem.uci.edu/content/illumina-guidelines>]

355. Patel RK, Jain M: **NGS QC Toolkit: a toolkit for quality control of next generation sequencing data.** *PloS one* 2012, **7**(2):e30619.
356. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD *et al.*: **SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.** *Journal of computational biology : a journal of computational molecular cell biology* 2012, **19**(5):455-477.
357. Nikolenko SI, Korobeynikov AI, Alekseyev MA: **BayesHammer: Bayesian clustering for error correction in single-cell sequencing.** *BMC Genomics* 2013, **14 Suppl 1**:S7.
358. Paszkiewicz K, Studholme DJ: **De novo assembly of short sequence reads.** *Briefings in bioinformatics* 2010, **11**(5):457-472.
359. Li H, Durbin R: **Fast and accurate long-read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2010, **26**(5):589-595.
360. **Picard command-line tools** [<http://broadinstitute.github.io/picard/>]
361. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I: **ABYSS: a parallel assembler for short read sequence data.** *Genome research* 2009, **19**(6):1117-1123.
362. Hernandez D, Tewhey R, Veyrieras JB, Farinelli L, Osteras M, Francois P, Schrenzel J: **De novo finished 2.8 Mbp Staphylococcus aureus genome assembly from 100 bp short and long range paired-end reads.** *Bioinformatics* 2014, **30**(1):40-49.
363. Zerbino DR, Birney E: **Velvet: algorithms for de novo short read assembly using de Bruijn graphs.** *Genome research* 2008, **18**(5):821-829.
364. Nagarajan N, Pop M: **Sequence assembly demystified.** *Nature reviews Genetics* 2013, **14**(3):157-167.
365. Gurevich A, Saveliev V, Vyahhi N, Tesler G: **QUAST: quality assessment tool for genome assemblies.** *Bioinformatics* 2013, **29**(8):1072-1075.
366. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ: **Prodigal: prokaryotic gene recognition and translation initiation site identification.** *BMC Bioinformatics* 2010, **11**:119.
367. Tourasse NJ, Helgason E, Okstad OA, Hegna IK, Kolsto AB: **The Bacillus cereus group: novel aspects of population structure and genome dynamics.** *J Appl Microbiol* 2006, **101**(3):579-593.
368. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: **BLAST+: architecture and applications.** *BMC Bioinformatics* 2009, **10**:421.
369. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J *et al.*: **Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.** *Mol Syst Biol* 2011, **7**:539.
370. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R: **Analysis Tool Web Services from the EMBL-EBI.** *Nucleic Acids Res* 2013, **41**(Web Server issue):W597-600.
371. Sievers F, Higgins DG: **Clustal Omega, accurate alignment of very large numbers of sequences.** *Methods Mol Biol* 2014, **1079**:105-116.
372. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**(22):4673-4680.
373. Kumar S, Stecher G, Tamura K: **MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.** *Mol Biol Evol* 2016, **33**(7):1870-1874.
374. Darriba D, Taboada GL, Doallo R, Posada D: **jModelTest 2: more models, new heuristics and parallel computing.** *Nat Methods* 2012, **9**(8):772.
375. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52**(5):696-704.
376. Akaike H: **Information theory and an extension of the Maximum likelihood Principal.** In: *2nd International Symposium on Information Theory, 1973: 1973: Akademiai Kiado; 1973: 267-281.*
377. Sugiura N: **Further analysis of the data by akaike's information criterion and the finite corrections: Further analysts of the data by akaike's.** *Communications in Statistics-Theory and Methods* 1978, **7**(1):13-26.

378. Hurvich CM, Tsai C-L: **Regression and time series model selection in small samples.** *Biometrika* 1989, **76**(2):297-307.
379. Schwarz G: **Estimating the dimension of a model.** *The annals of statistics* 1978, **6**(2):461-464.
380. Tavaré S: **Some probabilistic and statistical problems in the analysis of DNA sequences.** *Lectures on mathematics in the life sciences* 1986, **17**(2):57-86.
381. Letunic I, Bork P: **Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees.** *Nucleic Acids Res* 2016, **44**(W1):W242-245.
382. Delcher AL, Phillippy A, Carlton J, Salzberg SL: **Fast algorithms for large-scale genome alignment and comparison.** *Nucleic Acids Res* 2002, **30**(11):2478-2483.
383. Delcher AL, Kasif S, Fleischmann RD, Peterson J, White O, Salzberg SL: **Alignment of whole genomes.** *Nucleic Acids Res* 1999, **27**(11):2369-2376.
384. Novichkov V, Kaznadzey A, Alexandrova N, Kaznadzey D: **NSimScan: DNA comparison tool with increased speed, sensitivity and accuracy.** *Bioinformatics* 2016, **32**(15):2380-2381.
385. Nelder JA, Wedderburn RW: **Generalized linear models.** *J R Stat Soc* 1972, **135**(3):370-384.
386. Pagani I, Liolios K, Jansson J, Chen IM, Smirnova T, Nosrat B, Markowitz VM, Kyrpides NC: **The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata.** *Nucleic Acids Res* 2012, **40**(Database issue):D571-579.
387. R Core Team: **R: A Language and Environment for Statistical Computing.** In. Vienna, Austria: R Foundation for Statistical Computing; 2016.
388. McFadden D: **Conditional logit analysis of qualitative choice behavior.** 1973.
389. Cox DR, Snell EJ: **Analysis of binary data**, vol. 32: CRC Press; 1989.
390. Nagelkerke NJ: **A note on a general definition of the coefficient of determination.** *Biometrika* 1991, **78**(3):691-692.
391. Aquino J: **descr: Descriptive Statistics.** . In.; 2016.
392. Sing T, Sander O, Beerenwinkel N, Lengauer T: **ROCR: visualizing classifier performance in R.** *Bioinformatics* 2005, **21**(20):3940-3941.
393. Brandes U, Eiglsperger M, Herman I, Himsolt M, Marshall MS: **GraphML progress report structural layer proposal.** In: *International Symposium on Graph Drawing: 2001*: Springer; 2001: 501-512.
394. Bray T, Paoli J, Sperberg-McQueen CM, Maler E, Yergeau F: **Extensible markup language (XML).** *World Wide Web Journal* 1997, **2**(4):27-66.
395. Bastian M, Heymann S, Jacomy M: **Gephi: an open source software for exploring and manipulating networks.** *Icwsn* 2009, **8**:361-362.
396. Miele V, Penel S, Duret L: **Ultra-fast sequence clustering from similarity networks with SiLiX.** *BMC Bioinformatics* 2011, **12**:116.
397. Miele V, Penel S, Daubin V, Picard F, Kahn D, Duret L: **High-quality sequence clustering guided by network topology and multiple alignment likelihood.** *Bioinformatics* 2012, **28**(8):1078-1085.
398. **Gnumeric Spreadsheet** [<http://www.gnumeric.org/>]
399. Saha R, Sproer C, Beck B, Bagley S: **Pseudomonas oleovorans subsp. lubricantis subsp. nov., and reclassification of Pseudomonas pseudoalcaligenes ATCC 17440T as later synonym of Pseudomonas oleovorans ATCC 8062 T.** *Curr Microbiol* 2010, **60**(4):294-300.
400. Lang E, Griese B, Sproer C, Schumann P, Steffen M, Verbarq S: **Characterization of 'Pseudomonas azelaica' DSM 9128, leading to emended descriptions of Pseudomonas citronellolis Seubert 1960 (Approved Lists 1980) and Pseudomonas nitroreducens lizuka and Komagata 1964 (Approved Lists 1980), including Pseudomonas multiresinivorans as its later heterotypic synonym.** *Int J Syst Evol Microbiol* 2007, **57**(Pt 4):878-882.
401. Wolterink AF, Jonker AB, Kengen SW, Stams AJ: **Pseudomonas chloritidismutans sp. nov., a non-denitrifying, chlorate-reducing bacterium.** *Int J Syst Evol Microbiol* 2002, **52**(Pt 6):2183-2190.
402. Baumann P, Bowditch R, L. B, Beaman B: **Taxonomy of Marine Pseudomonas Species: P. stanieri sp. nov.; P. perfectomarina sp. nov., nom. rev.; P. nautica: and P. doudoroffii.** *International Journal of Systematic and Evolutionary Microbiology* 1983, **33**(4):857-865.

403. Cladera AM, Garcia-Valdes E, Lalucat J: **Genotype versus phenotype in the circumscription of bacterial species: the case of *Pseudomonas stutzeri* and *Pseudomonas chloritidismutans***. *Arch Microbiol* 2006, **184**(6):353-361.
404. Döhler K, Huss VA, Zumft WG: **Transfer of *Pseudomonas perfectomarina* Baumann, Bowditch, Baumann, and Beaman 1983 to *Pseudomonas stutzeri* (Lehmann and Neumann 1896) Sijderius 1946**. *International Journal of Systematic and Evolutionary Microbiology* 1987, **37**(1):1-3.
405. Rosselló-Mora RA, Garcia-Valdes E, Lalucat J: **Taxonomic relationship between *Pseudomonas perfectomarina* ZoBell and *Pseudomonas stutzeri***. *International Journal of Systematic and Evolutionary Microbiology* 1993, **43**(4):852-854.
406. Nakhimovskaya MI: ***Pseudomonas aurantiaca* n. sp.** *Mikrobiologiya* 1948, **17**:58-65.
407. Kluyver AJ: ***Pseudomonas aureofaciens* nov. spec. and its pigments**. *J Bacteriol* 1956, **72**(3):406-411.
408. Guignard L, Sauvageau C: **Sur un nouveau Microbe chromogène le *Bacillus Chlororaphis***. *Comptes Rendus des Séances de la Société de Biologie (Paris)*, 1894, **Séries 10**(1):841-843.
409. Burr SE, Gobeli S, Kuhnert P, Goldschmidt-Clermont E, Frey J: ***Pseudomonas chlororaphis* subsp. *piscium* subsp. nov., isolated from freshwater fish**. *Int J Syst Evol Microbiol* 2010, **60**(Pt 12):2753-2757.
410. Staley J, Krieg NR: **Bacterial classification I. Classification of procaryotic organisms: an overview**. In: *Bergey's Manual of Systematic Bacteriology*. Edited by Krieg NR, Holt JG, vol. 1: Williams & Wilkins Co, Baltimore; 1984: 1-4.
411. Schaad NW, Postnikova E, Sechler A, Claflin LE, Vidaver AK, Jones JB, Agarkova I, Ignatov A, Dickstein E, Ramundo BA: **Reclassification of subspecies of *Acidovorax avenae* as *A. Avenae* (Manns 1905) emend., *A. cattleyae* (Pavarino, 1911) comb. nov., *A. citrulli* Schaad et al., 1978) comb. nov., and proposal of *A. oryzae* sp. nov.** *Syst Appl Microbiol* 2008, **31**(6-8):434-446.
412. Li X, Tambong J, Yuan KX, Chen W, Xu H, Levesque CA, De Boer SH: **Re-classification of *Clavibacter michiganensis* subspecies on the basis of whole-genome and multi-locus sequence analyses**. *Int J Syst Evol Microbiol* 2018, **68**(1):234-240.
413. Liu Y, Lai Q, Shao Z: **Genome analysis-based reclassification of *Bacillus weihenstephanensis* as a later heterotypic synonym of *Bacillus mycoides***. *Int J Syst Evol Microbiol* 2018, **68**(1):106-112.
414. Perego M, Hoch JA: **Commingleing regulatory systems following acquisition of virulence plasmids by *Bacillus anthracis***. *Trends Microbiol* 2008, **16**(5):215-221.
415. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S et al: **High-quality draft assemblies of mammalian genomes from massively parallel sequence data**. *Proc Natl Acad Sci U S A* 2011, **108**(4):1513-1518.
416. Peng Z, Zhao Z, Nath N, Froula JL, Clum A, Zhang T, Cheng JF, Copeland AC, Pennacchio LA, Chen F: **Generation of long insert pairs using a Cre-LoxP Inverse PCR approach**. *PLoS one* 2012, **7**(1):e29437.
417. Koren S, Harhay GP, Smith TP, Bono JL, Harhay DM, McVey SD, Radune D, Bergman NH, Phillippy AM: **Reducing assembly complexity of microbial genomes with single-molecule sequencing**. *Genome Biol* 2013, **14**(9):R101.
418. Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED et al: **Hybrid error correction and de novo assembly of single-molecule sequencing reads**. *Nat Biotechnol* 2012, **30**(7):693-700.
419. Schroder J, Schroder H, Puglisi SJ, Sinha R, Schmidt B: **SHREC: a short-read error correction method**. *Bioinformatics* 2009, **25**(17):2157-2163.
420. Kelley DR, Schatz MC, Salzberg SL: **Quake: quality-aware detection and correction of sequencing errors**. *Genome Biol* 2010, **11**(11):R116.
421. Rodriguez-R LM, Konstantinidis KT: **The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes**. *PeerJ Preprints* 2016.
422. Shin SK, Hwang CY, Cho YJ, Yi H: **Reclassification of *Serpens flexibilis* Hespell 1977 as *Pseudomonas flexibilis* comb. nov., with *Pseudomonas tuomuerensis* Xin et al. 2009 as a later heterotypic synonym**. *Syst Appl Microbiol* 2015, **38**(8):563-566.
423. Rossello R, García-Valdés E, Lalucat J, Ursing J: **Genotypic and phenotypic diversity of *Pseudomonas stutzeri***. *Systematic and applied microbiology* 1991, **14**(2):150-157.

424. Lalucat J, Bennasar A, Bosch R, Garcia-Valdes E, Palleroni NJ: **Biology of *Pseudomonas stutzeri***. *Microbiol Mol Biol Rev* 2006, **70**(2):510-547.
425. Ursing JB, Rossellö-Mora RA, Garcia-Valdes E, Lalucat J: **Taxonomic note: a pragmatic approach to the nomenclature of phenotypically similar genomic groups**. *International Journal of Systematic and Evolutionary Microbiology* 1995, **45**(3):604-604.
426. Brenner DJ, Staley JT, Krieg NR: **Classification of procaryotic organisms and the concept of bacterial speciation**. In: *Bergey's Manual of Systematic Bacteriology*. Springer; 2005: 27-32.
427. Bennasar A, Rossello-Mora R, Lalucat J, Moore ER: **16S rRNA gene sequence analysis relative to genomovars of *Pseudomonas stutzeri* and proposal of *Pseudomonas balearica* sp. nov.** *Int J Syst Bacteriol* 1996, **46**(1):200-205.
428. Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P: **Toward automatic reconstruction of a highly resolved tree of life**. *Science* 2006, **311**(5765):1283-1287.
429. Creevey CJ, Doerks T, Fitzpatrick DA, Raes J, Bork P: **Universally distributed single-copy genes indicate a constant rate of horizontal transfer**. *PloS one* 2011, **6**(8):e22099.
430. Mende DR, Sunagawa S, Zeller G, Bork P: **Accurate and universal delineation of prokaryotic species**. *Nat Methods* 2013, **10**(9):881-884.
431. Gardan L, Shafik H, Belouin S, Broch R, Grimont F, Grimont PA: **DNA relatedness among the pathovars of *Pseudomonas syringae* and description of *Pseudomonas tremae* sp. nov. and *Pseudomonas cannabina* sp. nov. (ex Sutic and Dowson 1959)**. *Int J Syst Bacteriol* 1999, **49 Pt 2**:469-478.
432. Hauser E, Kampfer P, Busse HJ: ***Pseudomonas psychrotolerans* sp. nov.** *Int J Syst Evol Microbiol* 2004, **54**(Pt 5):1633-1637.
433. Park YD, Lee HB, Yi H, Kim Y, Bae KS, Choi JE, Jung HS, Chun J: ***Pseudomonas panacis* sp. nov., isolated from the surface of rusty roots of Korean ginseng**. *Int J Syst Evol Microbiol* 2005, **55**(Pt 4):1721-1724.
434. Wu M, Wen J, Chang M, Yang G, Zhou S: ***Pseudomonas sihuiensis* sp. nov., isolated from a forest soil in South China**. *Antonie Van Leeuwenhoek* 2014, **105**(4):781-790.
435. Miyajima K, Tanii A, Akita T: ***Pseudomonas fuscovaginae* sp. nov., nom. rev.** *International Journal of Systematic and Evolutionary Microbiology* 1983, **33**(3):656-657.
436. Xie F, Ma H, Quan S, Liu D, Chen G, Chao Y, Qian S: ***Pseudomonas kunmingensis* sp. nov., an exopolysaccharide-producing bacterium isolated from a phosphate mine**. *Int J Syst Evol Microbiol* 2014, **64**(Pt 2):559-564.
437. Feng Z, Zhang J, Huang X, Zhang J, Chen M, Li S: ***Pseudomonas zeshuii* sp. nov., isolated from herbicide-contaminated soil**. *Int J Syst Evol Microbiol* 2012, **62**(Pt 11):2608-2612.
438. Liu Y, Lai Q, Goker M, Meier-Kolthoff JP, Wang M, Sun Y, Wang L, Shao Z: **Genomic insights into the taxonomic status of the *Bacillus cereus* group**. *Sci Rep* 2015, **5**:14082.
439. Liu Y, Du J, Lai Q, Zeng R, Ye D, Xu J, Shao Z: **Proposal of nine novel species of the *Bacillus cereus* group**. *Int J Syst Evol Microbiol* 2017, **67**(8):2499-2508.
440. Beecher DJ, Wong AC: **Cooperative, synergistic and antagonistic haemolytic interactions between haemolysin BL, phosphatidylcholine phospholipase C and sphingomyelinase from *Bacillus cereus***. *Microbiology* 2000, **146 Pt 12**:3033-3039.
441. Williams LD, Burdock GA, Jimenez G, Castillo M: **Literature review on the safety of Toyocerin, a non-toxicogenic and non-pathogenic *Bacillus cereus* var. toyoi preparation**. *Regul Toxicol Pharmacol* 2009, **55**(2):236-246.
442. Andreeva ZI, Nesterenko VF, Yurkov IS, Budarina ZI, Sineva EV, Solonin AS: **Purification and cytotoxic properties of *Bacillus cereus* hemolysin II**. *Protein Expr Purif* 2006, **47**(1):186-193.
443. Cadot C, Tran SL, Vignaud ML, De Buyser ML, Kolsto AB, Brisabois A, Nguyen-The C, Lereclus D, Guinebretiere MH, Ramarao N: **InhA1, NprA, and HlyII as candidates for markers to differentiate pathogenic from nonpathogenic *Bacillus cereus* strains**. *J Clin Microbiol* 2010, **48**(4):1358-1365.
444. Jessberger N, Dietrich R, Bock S, Didier A, Martlbauer E: ***Bacillus cereus* enterotoxins act as major virulence factors and exhibit distinct cytotoxicity to different human cell lines**. *Toxicon* 2014, **77**:49-57.

445. Jessberger N, Krey VM, Rademacher C, Bohm ME, Mohr AK, Ehling-Schulz M, Scherer S, Martlbauer E: **From genome to toxicity: a combinatory approach highlights the complexity of enterotoxin production in *Bacillus cereus***. *Front Microbiol* 2015, **6**:560.
446. Ausschuss für Biologische Arbeitsstoffe: **TRBA 466 "Einstufung von Prokaryonten (Bacteria und Archaea) in Risikogruppen"** *Gemeinsames Ministerialblatt (GMBI)* 2015, **Nr. 46-50**.
447. Okinaka RT, Cloud K, Hampton O, Hoffmaster AR, Hill KK, Keim P, Koehler TM, Lamke G, Kumano S, Mahillon J *et al*: **Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes**. *J Bacteriol* 1999, **181**(20):6509-6515.
448. Berry C, O'Neil S, Ben-Dov E, Jones AF, Murphy L, Quail MA, Holden MT, Harris D, Zaritsky A, Parkhill J: **Complete sequence and organization of pBtoxis, the toxin-coding plasmid of *Bacillus thuringiensis* subsp. israelensis**. *Appl Environ Microbiol* 2002, **68**(10):5082-5095.
449. Francis KP, Mayr R, von Stetten F, Stewart GS, Scherer S: **Discrimination of psychrotrophic and mesophilic strains of the *Bacillus cereus* group by PCR targeting of major cold shock protein genes**. *Appl Environ Microbiol* 1998, **64**(9):3525-3529.
450. Pruss BM, Francis KP, von Stetten F, Scherer S: **Correlation of 16S ribosomal DNA signature sequences with temperature-dependent growth rates of mesophilic and psychrotolerant strains of the *Bacillus cereus* group**. *J Bacteriol* 1999, **181**(8):2624-2630.
451. Fiedoruk K, Drewnowska JM, Daniluk T, Leszczynska K, Iwaniuk P, Swiecicka I: **Ribosomal background of the *Bacillus cereus* group thermotypes**. *Sci Rep* 2017, **7**:46430.
452. Klee SR, Brzuszkiewicz EB, Nattermann H, Bruggemann H, Dupke S, Wollherr A, Franz T, Pauli G, Appel B, Liebl W *et al*: **The genome of a *Bacillus* isolate causing anthrax in chimpanzees combines chromosomal properties of *B. cereus* with *B. anthracis* virulence plasmids**. *PLoS one* 2010, **5**(7):e10986.
453. Okinaka R, Pearson T, Keim P: **Anthrax, but not *Bacillus anthracis*?** *PLoS Pathog* 2006, **2**(11):e122.

## VIII. ABBILDUNGSVERZEICHNIS

<b>Abbildung 1</b> Insert-Längenverteilungen nach der zweiten Längenselektion .....	67
<b>Abbildung 2</b> Gemessene und tatsächliche $\emptyset$ Insert-Längen aller IS1-3 Bibliotheken .....	68
<b>Abbildung 3</b> Einfluss der $\emptyset$ Insert-Länge und des GC-Gehaltes auf die Qualität der Sequenzierung .....	70
<b>Abbildung 4</b> Wechselwirkung zwischen Insert-Länge, GC-Gehalt und der Qualität der Sequenzierung einer Bibliothek .....	71
<b>Abbildung 5</b> Auswirkung der Filter-Einstellungen auf die durchschnittliche Insert-Länge einer Bibliothek .....	72
<b>Abbildung 6</b> Logit-Modelle der genomischen Ähnlichkeitsmaße ANIb und ANIm .....	77
<b>Abbildung 7</b> <i>Pseudomonas</i> MLSA ML-Teilbaum 1 .....	88
<b>Abbildung 8</b> <i>Pseudomonas</i> MLSA ML-Teilbaum 2 .....	89
<b>Abbildung 9</b> <i>Pseudomonas</i> MLSA ML-Teilbaum 3 .....	90
<b>Abbildung 10</b> Index-spezifische Spezies-Cluster ( <i>P. chlororaphis</i> ) .....	94
<b>Abbildung 11</b> Kodierung Index-spezifischer Spezies-Cluster in Blöcken .....	95
<b>Abbildung 12</b> <i>B. cereus s.l.</i> MLSA ML-Teilbaum 1 .....	104
<b>Abbildung 13</b> <i>B. cereus s.l.</i> MLSA ML-Teilbaum 2 .....	105
<b>Abbildung 14</b> Verteilung von <i>B. cereus s.l.</i> Virulenzfaktoren .....	110
<b>Abbildung 15</b> Phylogenie des Substrat-Bindeprotein Genes des Bikarbonat-Transporters .....	112
<b>Abbildung 16</b> Verteilung von Bt-Toxinen (Teilbaum 1) .....	114
<b>Abbildung 17</b> Verteilung von Bt-Toxinen (Teilbaum 2) .....	115
<b>Abbildung 18</b> Phylogenien basierend auf den Genen <i>cytK</i> und <i>hlyII</i> bzw. <i>hlyIII</i> und <i>hlyIIIIH</i> .....	118
<b>Abbildung 19</b> Gegenüberstellung der <i>B. cereus s.l.</i> MLSA und <i>hlyA</i> ML-Phylogenien .....	120
<b>Abbildung 20</b> Phylogenie der Gene aus der homologen Familie der Immulinhibitoren .....	122
<b>Abbildung 21</b> Einfluss der Read-Länge auf die Leistungsfähigkeit von Assemblierungen .....	131
<b>Abbildung 22</b> Abhängigkeiten der Index-spezifischen Spezies-Cluster Vorhersage ( <i>P. chlororaphis</i> ) .....	136
<b>Abbildung 23</b> Vorhergesagte Genomospezies-Cluster im <i>B. cereus s.l.</i> Datensatz .....	145

## IX. TABELLENVERZEICHNIS

<b>Tabelle 1</b>	Verwendete bakterielle Stämme (Optimierung Probenvorbereitung) .....	28
<b>Tabelle 2</b>	Stämme, die in dieser Arbeit sequenziert wurden .....	29
<b>Tabelle 3</b>	Kategorien der Sequenzierbibliotheken .....	38
<b>Tabelle 4</b>	Verwendetes MLSA Schema für die Gattung <i>Pseudomonas</i> .....	53
<b>Tabelle 5</b>	Verwendetes MLSA Schema für die <i>B. cereus s. l.</i> Gruppe.....	54
<b>Tabelle 6</b>	Verwendete Query-Referenzsequenzen (MLSA Screening).....	55
<b>Tabelle 7</b>	Angewendete Spezies-Grenzwerte zur Ermittlung von Spezies-Clustern.....	62
<b>Tabelle 8</b>	Relative Assemblierungs-Werte der verwendeten Assembler .....	74
<b>Tabelle 9</b>	Relative Assemblierungs-Werte untersuchter durchschnittlicher Insert-Längen (SPAdes).....	75
<b>Tabelle 10</b>	Relative Assemblierungs-Werte untersuchter Read-Längen (SPAdes).....	76
<b>Tabelle 11</b>	Relative Assemblierungs-Werte untersuchter Sequenzertiefen (SPAdes).....	76
<b>Tabelle 12</b>	Regressionskoeffizienten der Logit-Modelle .....	78
<b>Tabelle 13</b>	Signifikanz der genomischen Ähnlichkeitsmaße der Logit-Modelle .....	78
<b>Tabelle 14</b>	Gesamtmodell- und Klassifikationsgüte der Logit-Modelle.....	78
<b>Tabelle 15</b>	Genomsequenzierungen und Assemblierungen von Stämmen der Gattung <i>Pseudomonas</i> .....	80
<b>Tabelle 16</b>	<i>Pseudomonas</i> Gruppen/Untergruppenzugehörigkeit neuer Typstämme .....	86
<b>Tabelle 17</b>	Vorgeschlagene Reklassifizierungen innerhalb der Gattung <i>Pseudomonas</i> .....	98
<b>Tabelle 18</b>	Typstammkonflikte innerhalb des <i>Pseudomonas</i> Datensatzes nach Abgleich mit der Benchmark-Taxonomie.....	99
<b>Tabelle 19</b>	Anzahl bekannter und potentieller neuer <i>Pseudomonas</i> Spezies .....	101
<b>Tabelle 20</b>	Genomsequenzierungen und Assemblierungen von Stämmen der <i>B. cereus s.l.</i> Gruppe .....	107
<b>Tabelle 21</b>	Anzahl bekannter und potentieller neuer <i>B. cereus s.l.</i> Spezies .....	107
<b>Tabelle 22</b>	Sequenzähnlichkeiten innerhalb der Proteinfamilie der Immunitätsinhibitoren.....	122
<b>Tabelle 23</b>	Virulenzfaktorprofile (inkl. Bt-Toxine) der valide beschriebenen <i>B. cereus s.l.</i> Arten .....	126
<b>Tabelle 24</b>	<i>Pseudomonas</i> und <i>B. cereus s.l.</i> Spezies Neubeschreibungen von 2015 bis 2017.....	134
<b>Tabelle 25</b>	Grundlegende und ableitbare Toxinovare .....	154
<b>Tabelle 26</b>	Fallbeispiele zum Toxinovar System.....	155
<b>Tabelle 27</b>	Bezogene NCBI Genomdaten für die Spezies-Grenzwertableitung.....	187
<b>Tabelle 28</b>	Verwendete <i>Pseudomonas</i> NCBI Genomdaten.....	190
<b>Tabelle 29</b>	Verworfenene <i>Pseudomonas</i> NCBI Genomdaten.....	199
<b>Tabelle 30</b>	Verwendete <i>Pseudomonas</i> NCBI Einzelgensequenzen .....	200
<b>Tabelle 31</b>	Verwendete <i>Bacillus cereus s. l.</i> NCBI Genomdaten .....	202
<b>Tabelle 32</b>	Verworfenene <i>Bacillus</i> NCBI Genomdaten .....	211
<b>Tabelle 33</b>	NCBI Protein Referenzsequenzen untersuchter <i>B. cereus s.l.</i> Virulenzfaktoren (ohne Bt-Toxine) ..	213
<b>Tabelle 34</b>	NCBI Protein Referenzsequenzen untersuchter Bt-Toxine .....	214
<b>Tabelle 35</b>	Angewendete RB:BR Volumenverhältnisse während der ersten Längenselektion .....	217

<b>Tabelle 36</b>	Zugrundeliegende Daten der Logit-Modelle .....	218
<b>Tabelle 37</b>	Durchschnittliche Insert-Länge sequenzierter Bibliotheken.....	224
<b>Tabelle 38</b>	Durchschnittliche Insert-Längen von Bibliotheken vor und nach der Qualitätskontrolle .....	225
<b>Tabelle 39</b>	Relative Assemblierungs-Werte untersuchter durchschnittlicher Insert-Längen (ABYSS) .....	226
<b>Tabelle 40</b>	Relative Assemblierungs-Werte untersuchter Read-Längen (ABYSS).....	226
<b>Tabelle 41</b>	Relative Assemblierungs-Werte untersuchter Sequenziertiefen (ABYSS) .....	227
<b>Tabelle 42</b>	Relative Assemblierungs-Werte untersuchter durchschnittlicher Insert-Längen (Velvet).....	227
<b>Tabelle 43</b>	Relative Assemblierungs-Werte untersuchter Read-Längen (Velvet) .....	228
<b>Tabelle 44</b>	Relative Assemblierungs-Werte untersuchter Sequenziertiefen (Velvet).....	228
<b>Tabelle 45</b>	Relative Assemblierungs-Werte untersuchter durchschnittlicher Insert-Längen (Edena) .....	229
<b>Tabelle 46</b>	Relative Assemblierungs-Werte untersuchter Read-Längen (Edena).....	229
<b>Tabelle 47</b>	Relative Assemblierungs-Werte untersuchter Sequenziertiefen (Edena) .....	230
<b>Tabelle 48</b>	<i>Pseudomonas</i> Taxonomie nach LPSN mit Stand vom April 2017 .....	231

## X. ABKÜRZUNGSVERZEICHNIS

### Einheiten

%	Prozent
% (v/v)	Volumenprozent
% (w/v)	Gewichtsprozent
°C	Grad Celsius
µl	Mikroliter
µm	Mikrometer
bp	Basenpaar(e)
C/B	<i>Cycles per Burst</i>
DF	<i>Duty Factor</i>
Du	<i>Duration</i>
g	Gramm bzw. relative Zentrifugalbeschleunigung (RZB)
h	Stunde(n)
KbE	koloniebildenden Einheit
kpb	Kilobasenpaare
l	Liter
M	Molar (mol/l)
m	Meter
Mbp	Megabasenpaare
mg	Milligramm
min	Minute(n)
ml	Milliliter
mm	Millimeter
mM	Millimolar
N	Normalität
ng	Nanogramm
nM	Nanomolar
nt	Nukleotid(e)
p.	<i>page</i>
PIP	<i>Peak Incident Power</i>
pMol	Picomolar
pp.	<i>pages</i>
rpm	<i>rounds per minute</i>
s	Sekunde(n)
µg	Mikrogramm

### Nukleobasen / Nukleotide

A	Adenin
ATP	Adenosintri-phosphat
C	Cytosin
cAMP	Cyclisches Adenosinmonophosphat
G	Guanin
T	Thymin

### Bakterielle Gattungen, Arten und Stämme

B.	<i>Bacillus</i>
Bce	<i>Bacillus cereus</i> F837/76
Bt	<i>B. thuringiensis</i>
Efa	<i>Enterococcus faecalis</i> OG1RF
Mlu	<i>Micrococcus luteus</i> NCTC 2665 <sup>T</sup>
P.	<i>Pseudomonas</i>

Pst *Pseudomonas stutzeri* ATCC 17588<sup>T</sup>  
 Sen *Salmonella* Typhimurium 14028S

### **Gene, Gencluster, Operons, Proteine und Proteinkomplexe**

AbrB *transition state regulatory protein (protein)*  
 Adk *adenylate kinase (protein)*  
*adk* *adenylate kinase (gene)*  
 CcpA *catabolite control protein (protein)*  
*ccpA* *catabolite control protein A (gene)*  
*cesA* Cereulid NRPS (Gen)  
 CesA Cereulid NRPS (Protein)  
*cesB* Cereulid NRPS (Gen)  
 CesB Cereulid NRPS (Protein)  
 CesC und CesD putativer ABC-Transporter (Proteine)  
 CesH Hydrolase/Acyltransferase (Protein)  
*cesHPTABCD* Cereulid-Synthetase-Gencluster (Gencluster)  
 CesP 4'-Phosphopantetheinyl Transferase (Protein)  
 CesT Typ II Thioesterase (Protein)  
 CodY globaler Regulator (Protein)  
 Cry *Crystal Protein (protein)*  
*cry* *Crystal Protein (gene)*  
 CspA *Cold Shock Protein A (protein)*  
*cya* Ödemfaktor (Gen)  
 Cyt *Cytolytic Protein (protein)*  
*cyt* *Cytolytic Protein (gene)*  
 CytK / HlyIV Cytotoxin K / Hämolyisin IV (Protein)  
*cytK / hlyIV* Cytotoxin K / Hämolyisin IV (Gen)  
 EF Ödemfaktor (Protein)  
 EntFM Enterotoxin FM (Protein)  
 ET Ödemtoxin (Proteinkomplex)  
 Fnr *Fumerate and Nitrate Reduction Regulator (protein)*  
 Fur Eisenaufnahme-Regulator (Protein)  
 GlpF *glycerol uptake facilitator (protein)*  
*glpF* *glycerol uptake facilitator (gene)*  
 GlpT *glycerol-3-phosphatetransporter (protein)*  
*glpT* *glycerol-3-phosphatetransporter (gene)*  
*gyrB* *DNA gyrase subunit B (gene)*  
 GyrB *DNA gyrase subunit B (protein)*  
 HBL Hämolyisin BL (Proteinkomplex)  
 HblA B Untereinheit des Hämolyisin BL (Protein)  
*hblA* B Untereinheit des Hämolyisin BL (Gen)  
 HBL<sub>a</sub> Hämolyisin BL homologer Komplex (Proteinkomplex)  
 HblA<sub>a</sub> B Untereinheit des Hämolyisin BL homologen Komplexes (Protein)  
*hblB* putatives Pseudogen nach *hblA* Duplikation (Gen)  
 HblC L<sub>2</sub> Untereinheit des Hämolyisins BL (Protein)  
 HblC<sub>a</sub> L<sub>2</sub> Untereinheit des Hämolyisins BL homologen Komplexes (Protein)  
*hblCDA<sub>a</sub>* Operon des Hämolyisin BL homologen Komplexes (Operon)  
*hblCDAB* Operon des Hämolyisin BL (Operon)  
 HblD L<sub>1</sub> Untereinheit des Hämolyisin BL (Protein)  
*hblD* L<sub>1</sub> Untereinheit des Hämolyisin BL (Gen)  
 HblD<sub>a</sub> L<sub>1</sub> Untereinheit des Hämolyisin BL homologen Komplexes (Protein)  
 HlyA Hämolyisin A (Protein)  
*hlyA* Hämolyisin A (Gen)  
 HlyI / CerO Hämolyisin I / Cereolysin O (Protein)  
 HlyII Hämolyisin II (Protein)

<i>hlyII</i>	Hämolysin II (Gen)
HlyIII	Hämolysin III (Protein)
<i>hlyIII</i>	Hämolysin III (Gen)
HlyIIIH	Hämolysin III Homolog (Protein)
<i>hlyIIIH</i>	Hämolysin III Homolog (Gen)
HlyIIR	Hämolysin II Repressor (Protein)
InhA1	Immulinhibitor A1 (Protein)
<i>inhA1</i>	Immulinhibitor A1 (Gen)
InhA2	Immulinhibitor A2 (Protein)
<i>inhA2</i>	Immulinhibitor A2 (Gen)
InhA3	Immulinhibitor A3 (Protein)
<i>inhA3</i>	Immulinhibitor A3 (Gen)
InhA4	Immulinhibitor A4 (Protein)
<i>inhA4</i>	Immulinhibitor A4 (Gen)
InhA5	Immulinhibitor A5 (Protein)
<i>inhA5</i>	Immulinhibitor A5 (Gen)
<i>lef</i>	Letalfaktor (Gen)
LF	Letalfaktor (Protein)
LT	Letaltoxin (Proteinkomplex)
MAP-Kinase	Mitogen-aktivierte Proteinkinase (Protein)
Nhe	nicht-hämolytisches Enterotoxin (Proteinkomplex)
Nhe <sub>a</sub>	nicht-hämolytisches Enterotoxin homologer Komplex (Proteinkomplex)
NheA	lytische Komponente des nicht-hämolytischen Enterotoxins (Protein)
Nhe <sub>a</sub>	nicht-hämolytisches Enterotoxin homologer Komplex (Proteinkomplex)
NheA <sub>a</sub>	lytische Komponente des nicht-hämolytischen Enterotoxin homologen Komplexes (Protein)
<i>nheABC</i>	Operon des nicht-hämolytischen Enterotoxins (Operon)
<i>nheABC<sub>a</sub></i>	Operon des nicht-hämolytischen Enterotoxin homologen Komplexes (Operon)
NheB	Bindeprotein des nicht-hämolytischen Enterotoxins (Protein)
NheB <sub>a</sub>	Bindeprotein des nicht-hämolytischen Enterotoxin homologen Komplexes (Protein)
NheC	Bindeprotein des nicht-hämolytischen Enterotoxins (Protein)
<i>nheC</i>	Bindeprotein des nicht-hämolytischen Enterotoxins (Gen)
NheC <sub>a</sub>	Bindeprotein des nicht-hämolytischen Enterotoxin homologen Komplexes (Protein)
PA	Protektives Antigen (Protein)
<i>pagA</i>	Protektives Antigen (Gen)
PanC	<i>pantoate-β-alanine ligase (protein)</i>
<i>panC</i>	<i>pantoate-β-alanine ligase (gene)</i>
PC-PLC	Phosphatidylcholin-spezifische Phospholipase C (Protein)
PI-PLC	Phosphatidylinositol-spezifische Phospholipase C (Protein)
<i>plcR</i>	Phospholipase C Regulator (Gen)
PlcR	Phospholipase C Regulator (Protein)
Pta	<i>phosphate acetyltransferase (protein)</i>
<i>pta</i>	<i>phosphate acetyltransferase (gene)</i>
PycA	<i>pyruvate carboxylase (protein)</i>
<i>pycA</i>	<i>pyruvate carboxylase (gene)</i>
ResDE	<i>Redox-Sensitive Two Component Signal Transduction System (proteins)</i>
<i>rpoB</i>	<i>RNA polymerase subunit B (gene)</i>
RpoB	<i>RNA polymerase subunit B (protein)</i>
<i>rpoD</i>	<i>RNA polymerase subunit D (gene)</i>
RpoD	<i>RNA polymerase subunit D (protein)</i>
<i>sip</i>	<i>Secreted Insecticidal Protein (gene)</i>
Sip	<i>Secreted Insecticidal Protein (protein)</i>
SM-PLC	Sphingomyelin-spezifische Phospholipase C (Protein)
Vip	<i>Vegetative Insecticidal Protein (protein)</i>
<i>vip</i>	<i>Vegetative Insecticidal Protein (gene)</i>

**Stammsammlungen**

ATCC	<i>American Type Culture Collection, Manassas, USA</i>
BC	<i>Bacterial Collection of the Institute of Microbiology, Messina, Italy</i>
BCRC	<i>Bioresource Collection and Research Center, Hsinchu, Taiwan</i>
BGSC	<i>Bacillus Genetic Stock Center, Columbus, USA</i>
BPIC	<i>Benaki Phytopathological Institute Collections of Phytopathogenic Fungi and Bacteria, Athens, Greece</i>
CCM	<i>Czech Collection of Microorganisms, Brno, Czech Republic</i>
CCTCC	<i>Chinese Centre for Type Cultures Collection, Wuhan, China</i>
CCUG	<i>Culture Collection University of Göteborg, Göteborg, Sweden</i>
CECT	<i>Colección Española de Cultivos Tipo, Valencia, Spain</i>
CFBP	<i>Collection Française de Bactéries associées aux plantes, Beaucauzé Cedex, France</i>
CFML	<i>Collection de la Faculté de Médecine de Lille, Lille Cedex, France</i>
CGMCC	<i>China General Microbiological Culture Collection Centre, Beijing, China</i>
CIP	<i>Collection de l'Institut Pasteur, Paris, France</i>
CVUAS	<i>Chemisches und Veterinäruntersuchungsamt Stuttgart, Stuttgart, Deutschland</i>
DSM	<i>Deutschen Sammlung von Mikroorganismen, Braunschweig, Deutschland</i>
FPC	<i>Fish Pathogen Collection, Tokyo, Japan</i>
IAMCC	<i>Institute of Applied Microbiology Culture Collection, Tokyo, Japan</i>
IBL	<i>Botanical Institute, Lisbon Faculty of Sciences, Lisbon, Portugal</i>
ICMP	<i>International Collection of Micro-organisms from Plants, Auckland, New Zealand</i>
IFO	<i>Institute for Fermentation, Culture Collection of Microorganisms, Osaka, Japan</i>
IMSNU	<i>Institute of Microbiology Seoul National University, Seoul, Republic of Korea</i>
INRA	<i>Microbiologie des Sols, Dijon Cedex, France</i>
JCM	<i>Japan Collection of Microorganisms, Tsukuba, Japan</i>
KACC	<i>Korean Agricultural Culture Collection, Suwon, Republic of Korea</i>
KCTC	<i>Korean Collection for Type Cultures, Jeongeup-si, Republic of Korea</i>
KMM	<i>Collection of Marine Microorganisms, Vladivostok, Russia</i>
LMG	<i>Collection of the Laboratorium voor Microbiologie en Microbiele Genetica, Gent, Belgium</i>
MCC	<i>Microbial Culture Collection, Pune, India</i>
MTCC	<i>Microbial Type Culture Collection, Chandigarh, India</i>
NBRC	<i>National Institute of Technology and Evaluation Biological Resource Center, Kisarazu-shi, Japan</i>
NCIMB	<i>National Collection of Industrial and Marine Bacteria, Aberdeen, UK</i>
NCPPB	<i>National Collection of Plant Pathogenic Bacteria, York, UK</i>
NCTC	<i>National Collection of Type Cultures, London, UK</i>
NRRL	<i>Northern Regional Research Center, Peoria, USA</i>
UCM	<i>Ukrainian Collection of Microorganisms, Kyiv, Ukraine</i>
WS	<i>Weihenstephaner Mikroorganismen-Sammlung, Freising-Weihenstephan, Deutschland</i>
WSBC	<i>Research collection of Bacillus cereus group species, Freising-Weihenstephan, Germany</i>

### **Phylogenomische Indices, Ähnlichkeitsmaße und Programme**

ANI	<i>Average Nucleotide Identity</i>
ANIb	<i>Average Nucleotide Identity via Standard Nucleotide BLAST</i>
ANIb <sub>1</sub>	<i>ANIb Implementierung 1</i>
ANIb <sub>2</sub>	<i>ANIb Implementierung 2</i>
ANIb <sub>JSpeciesWS</sub>	<i>ANIb mittels des JSpecies Web Servers</i>
ANIm	<i>Average Nucleotide Identity via MUMmer</i>
ANIm <sub>JSpeciesWS</sub>	<i>ANIm mittels des JSpecies Web Servers</i>
FastANI	<i>Fast Average Nucleotide Identity</i>
gANI	<i>genome-wide Average Nucleotide Identity</i>
GBDP	<i>Genome-Blast-Distance-Phylogeny</i>
GGDC	<i>Genome-to-Genome Distance Calculator</i>
MiSI	<i>Microbial Species Identifier</i>
OAT	<i>Orthologous Average Nucleotide Identity Tool</i>
OrthoANI	<i>Average Nucleotide Identity by Orthology</i>
OrthoANLu	<i>OrthoANI via USEARCH</i>

### **Weitere Programme und Web Services**

BLAST	<i>Basic Local Alignment Search Tool</i>
blastn	<i>Standard Nucleotide BLAST</i>
blastp	<i>Protein-Protein BLAST</i>
HiFiX	<i>High Fidelity Clustering of Sequences</i>
iTOL	<i>Interactive Tree Of Life</i>
PHYLP	<i>Phylogeny Inference Package</i>
QUAST	<i>Quality Assessment Tool for Genome Assemblies</i>
SiLiX	<i>Single Linkage Clustering of Sequences</i>
specl	<i>Species Identification Tool</i>

### **Biologische Datenbanken**

GenBank	<i>Genetic Sequence Database</i>
GOLD	<i>Genomes OnLine Database</i>
LPSN	<i>List of Prokaryotic names with Standing in Nomenclature</i>
NCBI	<i>National Center for Biotechnology Information</i>
RefSeq	<i>NCBI Reference Sequence Database</i>
SRA	<i>Sequencing Read Archive</i>
WGS	<i>Whole Genome Shotgun</i>

### **Weitere Abkürzungen**

AF	<i>Alignment Fraction</i>
AFLP	<i>Amplifizierter Fragmentlängen-Polymorphismus</i>
AIC	<i>Akaike Informationskriterium (Akaike information criterion)</i>
AIC <sub>c</sub>	<i>korrigiertes Akaike Informationskriterium (corrected AIC)</i>
Ala	<i>Alanin (Aminosäure)</i>
Apr.	<i>April</i>
AUC	<i>Area under curve</i>
BBH	<i>Bidirektional-Bester-Hit</i>
BIC	<i>Bayesianisches Informationskriterium (Bayesian information criterion)</i>
BR	<i>Agencourt AMPure® XP magnetic bead Reagent</i>
Ca	<i>Calcium (Element)</i>
CASO-Agar	<i>Caseinpepton-Sojamehlpepton-Agar</i>
CASO-Bouillon	<i>Caseinpepton-Sojamehlpepton-Bouillon</i>
CDS	<i>proteinkodierende Sequenz(en) (coding sequence)</i>
CTAB	<i>Hexadecyltrimethylammoniumbromid</i>
dBG	<i>de Bruijn Graph</i>
dDDH	<i>digitaler DNA-DNA Hybridisierungswert</i>
DDH	<i>DNA-DNA Hybridisierung</i>
DNA	<i>Desoxyribonukleinsäure (deoxyribonucleic acid)</i>
EDTA	<i>Ethylendiamin-tetraessigsäure</i>
<i>et al.</i>	<i>et alii</i>
<i>etc.</i>	<i>et cetera</i>
<i>exp.</i>	<i>experimentell</i>
<i>Feb.</i>	<i>Februar</i>
FTIR Spectroscopy	<i>Fourier Transform Infrared Spectroscopy</i>
GAGE	<i>Genome Assembly Gold-Standard Evaluations</i>
GAGE-B	<i>Genome Assembly Gold-standard Evaluation for Bacteria</i>
GEBA	<i>Genomic Encyclopedia of Bacteria and Archaea</i>
GTR	<i>General Time Reversible</i>
H <sub>2</sub> O <sub>dest.</sub>	<i>destilliertes Wasser</i>
HCl	<i>Salzsäure</i>
ICR-FT MS	<i>Ion Cyclotron Resonance-Fourier Transform Mass Spectrometry</i>
ID	<i>Identifizierer (identifier)</i>
inkl.	<i>inklusive</i>
IS1	<i>Sequenzierbibliothek der Kategorie IS1</i>
IS2	<i>Sequenzierbibliothek der Kategorie IS2</i>

IS3	Sequenzierbibliothek der Kategorie IS3
IS4	Sequenzierbibliothek der Kategorie IS4
iT	<i>invalid type strain</i>
JC	Jukes-Cantor
K	Kalium (Element)
Leu	Leucin (Aminosäure)
MALDI-TOF MS	<i>Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight Mass Spectrometry</i>
MEE	Multilocus-Enzym-Elektrophorese
ML	Maximum-Likelihood
MLSA	Multilocus-Sequenz-Analyse
MLST	Multilocus-Sequenz-Typisierung
NaCl	Natriumchlorid
NaOH	Natronlauge
NGS	<i>Next-Generation-Sequencing</i>
NRPS	nichtribosomale Peptidsynthetase
OLC	<i>Overlap-Layout-Consensus</i>
PacBio	<i>Pacific Bioscience</i>
PCR	Polymerase-Kettenreaktion ( <i>polymerase chain reaction</i> )
qPCR	quantitative PCR
RB	<i>Resuspension Buffer</i>
rk	reverse komplementierte Sequenz
RNA	Ribonukleinsäure ( <i>ribonucleic acid</i> )
ROC	<i>Receiver-Operating-Characteristic</i>
rRNA	ribosomale RNA
RT	Raumtemperatur
S	Singleton
<i>s.l.</i>	<i>sensu lato</i>
<i>s.s.</i>	<i>sensu stricto</i>
SAM	<i>Sequence Alignment/Map</i>
SDS	Natriumlaurylsulfat ( <i>sodium dodecyl sulfate</i> )
SMRT	<i>Single Molecule Real Time</i>
sp.	<i>species</i>
sp. nov.	<i>species nova</i>
spp.	<i>species pluralis</i>
ssp.	<i>subspecies</i>
T	Typstamm
TE-Puffer	TRIS-EDTA-Puffer
TGS	<i>Third-Generation-Sequencing</i>
TRIS	Tris(hydroxymethyl)-aminomethan
TS	Sequenzierbibliothek der Kategorie TS
ÜNK	Übernachtkultur
Val	Valin (Aminosäure)
vs.	<i>versus</i>
WS	<i>Web Service</i> bzw. <i>Web Server</i>
XML	Extensible Markup Language
z.B.	zum Beispiel
Zn	Zink (Element)

## XI. ANHANG

**Tabelle 27 Bezogene NCBI Genomdaten für die Spezies-Grenzwertableitung**

<b>Stammbezeichner<sub>NCBI</sub></b>	<b>WGS Kennung</b>	<b>RefSeq Sequenzen</b>
<i>Bacillus cereus</i> ATCC 10987		NC_003909.8, NC_005707.1
<i>Bacillus cereus</i> ATCC 14579 (T)		NC_004721.2, NC_004722.1
<i>Bacteroides cellulosilyticus</i> DSM 14838 (T)	ACCH01 WGS_SCAFLD	
<i>Bacteroides dorei</i> DSM 17855 (T)	ABWZ01	
<i>Bacteroides fingoldii</i> DSM 17565 (T)	ABXI02	
<i>Bacteroides fragilis</i> NCTC 9343 (T)		NC_003228.3, NC_006873.1
<i>Bacteroides fragilis</i> YCH46		NC_006297.1, NC_006347.1
<i>Bacteroides intestinalis</i> DSM 17393 (T)	ABJL02	
<i>Bacteroides ovatus</i> ATCC 8483 (T)	AAXF02	
<i>Bacteroides thetaiotaomicron</i> VPI-5482 (T)		NC_004663.1, NC_004703.1
<i>Bacteroides uniformis</i> ATCC 8492 (T)	AAYH02	
<i>Bacteroides vulgatus</i> ATCC 8482 (T)		NC_009614.1
<i>Bifidobacterium angulatum</i> DSM 20098 (T)	JDTY01	
<i>Bifidobacterium breve</i> DSM 20213 (T)	JDUD01	
<i>Bifidobacterium catenulatum</i> DSM 16992 (T)	JDTP01	
<i>Bifidobacterium gallicum</i> DSM 20093 (T)	JDTW01	
<i>Bifidobacterium pseudocatenulatum</i> DSM 20438 (T)	JDUK01	
<i>Brucella canis</i> ATCC 23365 (T)		NC_010103.1, NC_010104.1
<i>Brucella melitensis</i> ATCC 23457		NC_012441.1, NC_012442.1
<i>Brucella ovis</i> ATCC 25840 (T)		NC_009504.1, NC_009505.1
<i>Brucella suis</i> 1330 (T)		NC_017250.1, NC_017251.1
<i>Brucella suis</i> ATCC 23445		NC_010167.1, NC_010169.1
<i>Burkholderia cenocepacia</i> AU 1054		NC_008060.1 bis NC_008062.1
<i>Burkholderia cenocepacia</i> HI2424		NC_008542.1 bis NC_008544.1
<i>Burkholderia cenocepacia</i> J2315 (T)		NC_011000.1 bis NC_011002.1
<i>Burkholderia lata</i> 383 (T)		NC_007509.1 bis NC_007511.1
<i>Burkholderia vietnamiensis</i> G4		NC_009226.1 bis NC_009230.1, NC_009254.1 bis NC_009256.1
<i>Burkholderia xenovorans</i> LB400 (T)		NC_007951.1 bis NC_007953.1
<i>Chloroflexus aggregans</i> DSM 9485 (T)		NC_011831.1
<i>Chloroflexus aurantiacus</i> J-10-fl (T)		NC_010175.1
<i>Clostridium hylemonae</i> DSM 15053 (T)	ABYI02	
<i>Clostridium scindens</i> ATCC 35704 (T)	ABFY02	
<i>Corynebacterium efficiens</i> YS-314 (T)		NC_004319.1, NC_004320.1, NC_004369.1
<i>Corynebacterium glutamicum</i> ATCC 13032 (T)		NC_003450.3

Stammbezeichner <sub>NCBI</sub>	WGS Kennung	RefSeq Sequenzen
<i>Corynebacterium striatum</i> ATCC 6940 (T)	ACGE01	
<i>Corynebacterium urealyticum</i> DSM 7109 (T)		NC_010545.1
<i>Cupriavidus metallidurans</i> CH34 (T)		NC_007971.2, NC_007972.2, NC_007973.1, NC_007974.2
<i>Deinococcus deserti</i> VCD115 (T)		NC_012526.1 bis NC_012529.1
<i>Deinococcus radiodurans</i> R1 (T)		NC_001263.1, NC_001264.1, NC_000958.1, NC_000959.1
<i>Escherichia coli</i> 042		NC_017626.1, NC_017627.1
<i>Escherichia coli</i> CFT073		NC_004431.1
<i>Escherichia coli</i> K-12 MG1655		NC_000913.3
<i>Escherichia coli</i> O127:H6 E2348/69		NC_011601.1 bis NC_011603.1
<i>Escherichia coli</i> O157:H7 EDL933		NC_002655.2, NC_007414.1
<i>Escherichia coli</i> O157:H7 Sakai		NC_002127.1, NC_002128.1, NC_002695.1
<i>Halorubrum lacusprofundi</i> ATCC 49239 (T)		NC_012028.1 bis NC_012030.1
<i>Haloterrigena turkmenica</i> DSM 5511 (T)		NC_013743.1 bis NC_013749.1
<i>Methanococcus maripaludis</i> C5		NC_009135.1, NC_009136.1
<i>Methanococcus maripaludis</i> C6		NC_009975.1
<i>Methanococcus maripaludis</i> C7		NC_009637.1
<i>Methanococcus vannielii</i> SB (T)		NC_009634.1
<i>Nitrosomonas europaea</i> ATCC 19718		NC_004757.1
<i>Nitrosomonas eutropha</i> C91		NC_008341.1, NC_008342.1, NC_008344.1
<i>Parabacteroides distasonis</i> ATCC 8503 (T)		NC_009615.1
<i>Parabacteroides johnsonii</i> DSM 18315 (T)	ABYH01 WGS_SCAFLD	
<i>Parabacteroides merdae</i> ATCC 43184 (T)	AAXE02	
<i>Pseudomonas aeruginosa</i> PAO1		NC_002516.2
<i>Pseudomonas fluorescens</i> Pf0-1		NC_007492.2
<i>Pseudomonas fluorescens</i> SBW25		NC_012660.1
<i>Pseudomonas protegens</i> Pf-5		NC_004129.6
<i>Pseudomonas syringae</i> B728a		NC_007005.1
<i>Pseudomonas syringae</i> DC3000		NC_004578.1, NC_004632.1, NC_004633.1
<i>Pyrobaculum aerophilum</i> IM2 (T)		NC_003364.1
<i>Pyrobaculum calidifontis</i> JCM 11548 (T)		NC_009073.1
<i>Pyrobaculum islandicum</i> DSM 4184 (T)		NC_008701.1
<i>Ralstonia eutropha</i> H16		NC_005241.1, NC_008313.1, NC_008314.1
<i>Salinibacter ruber</i> DSM 13855 (T)		NC_007677.1, NC_007678.1
<i>Salinibacter ruber</i> M8		NC_014026.1, NC_014028.1, NC_014030.1, NC_014032.1, NC_014157.1
<i>Selenomonas noxia</i> ATCC 43541 (T)	ACKT01	
<i>Selenomonas sputigena</i> ATCC 35185 (T)		NC_015437.1
<i>Shewanella amazonensis</i> SB2B (T)		NC_008700.1
<i>Shewanella baltica</i> OS155		NC_009035.1 bis NC_009038.1, NC_009052.1
<i>Shewanella baltica</i> OS185		NC_009661.1, NC_009665.1
<i>Shewanella baltica</i> OS195		NC_009997.1 bis NC_010000.1

<b>Stammbezeichner<sub>NCBI</sub></b>	<b>WGS Kennung</b>	<b>RefSeq Sequenzen</b>
<i>Shewanella baltica</i> OS223		NC_011663.1 bis NC_011665.1, NC_011668.1
<i>Shewanella halifaxensis</i> HAW-EB4 (T)		NC_010334.1
<i>Shewanella oneidensis</i> MR-1 (T)		NC_004347.2, NC_004349.1
<i>Shewanella pealeana</i> ATCC 700345 (T)		NC_009901.1
<i>Shewanella putrefaciens</i> CN-32		NC_009438.1
<i>Shewanella sediminis</i> HAW-EB3 (T)		NC_009831.1
<i>Shewanella woodyi</i> ATCC 51908 (T)		NC_010506.1
<i>Shigella flexneri</i> 2a 2457 (T)		NC_004741.1
<i>Shigella sonnei</i> 53G		NC_016822.1 bis NC_016824.1, NC_016833.1, NC_016834.1
<i>Streptococcus agalactiae</i> 2603V/R		NC_004116.1
<i>Streptococcus agalactiae</i> A909		NC_007432.1
<i>Streptococcus agalactiae</i> NEM316		NC_004368.1
<i>Sulfolobus acidocaldarius</i> DSM 639 (T)		NC_007181.1
<i>Sulfolobus solfataricus</i> P2		NC_002754.1
<i>Thermotoga maritima</i> MSB8 (T)		NC_000853.1
<i>Thermotoga neapolitana</i> DSM 4359 (T)		NC_011978.1
<i>Thermotoga petrophila</i> RKU-1 (T)		NC_009486.1
<i>Yersinia aldovae</i> ATCC 35236 (T)	ACCB01	
<i>Yersinia bercovieri</i> ATCC 43970 (T)	AALC02	
<i>Yersinia frederiksenii</i> ATCC 33641 (T)	JPPS01	
<i>Yersinia intermedia</i> ATCC 29909 (T)	AALF02	
<i>Yersinia mollaretii</i> ATCC 43969 (T)	AALD02	
<i>Yersinia rohdei</i> ATCC 43380 (T)	ACCD01	

**Stammbezeichner<sub>NCBI</sub>**, Bezeichnung des Stammes, wobei Gattungs- und Artzugehörigkeit sowie die Stammkennung vom NCBI übernommen wurden

**WGS Kennung**, Kennung des WGS Projektes bestehend aus einem 4 buchstabiligen Präfix (Projekt) gefolgt von 2 Ziffern (Versionsnummer)

**RefSeq Sequenzen**, Kennungen der NCBI Referenzgenomsequenzen

(T), anerkannter Typstamm nach LPSN [67]

WGS\_SCAFLD, Bezug der WGS Scaffold Sequenzen

Tabelle 28 Verwendete *Pseudomonas* NCBI Genomdaten

ID	Stammbezeichner <sub>NCBI</sub>	WGS Kennung	RefSeq/GenBank Sequenzen	Datum Bezug
P101	<i>P. abietaniphila</i> ATCC 700689 (T)	FNCO01		Apr. 2017
P102	<i>P. abietaniphila</i> KF717	BBQR01		Feb. 2015
P103	<i>P. aeruginosa</i> 19BR		NZ_AFXJ01000001.1	Feb. 2015
P104	<i>P. aeruginosa</i> 213BR		NZ_AFXK01000001.1	Feb. 2015
P105	<i>P. aeruginosa</i> 39016		NZ_CM001020.1	Feb. 2015
P106	<i>P. aeruginosa</i> B136-33		NC_020912.1	Feb. 2015
P107	<i>P. aeruginosa</i> c7447m		NC_022360.1	Feb. 2015
P108	<i>P. aeruginosa</i> DK2		NC_018080.1	Feb. 2015
P109	<i>P. aeruginosa</i> LES400		NZ_CP006982.1	Feb. 2015
P110	<i>P. aeruginosa</i> LES431		NC_023066.1	Feb. 2015
P111	<i>P. aeruginosa</i> LESB58		NC_011770.1	Feb. 2015
P112	<i>P. aeruginosa</i> LESB65		NZ_CP006983.1	Feb. 2015
P113	<i>P. aeruginosa</i> LESlike1		CP006984.1	Feb. 2015
P114	<i>P. aeruginosa</i> LESlike4		NZ_CP006985.1	Feb. 2015
P115	<i>P. aeruginosa</i> LESlike5		NZ_CP006980.1	Feb. 2015
P116	<i>P. aeruginosa</i> LESlike7		NZ_CP006981.1	Feb. 2015
P117	<i>P. aeruginosa</i> M18		NC_017548.1	Feb. 2015
P118	<i>P. aeruginosa</i> MTB-1		NC_023019.1	Feb. 2015
P119	<i>P. aeruginosa</i> NCGM2.S1		NC_017549.1	Feb. 2015
P120	<i>P. aeruginosa</i> PA1		NC_022808.1	Feb. 2015
P121	<i>P. aeruginosa</i> PA1R		NC_022806.1	Feb. 2015
P122	<i>P. aeruginosa</i> PA7		NC_009656.1	Feb. 2015
P123	<i>P. aeruginosa</i> PA96		CP007224.1	Feb. 2015
P124	<i>P. aeruginosa</i> PACS2		NZ_AAQW01000001.1	Feb. 2015
P125	<i>P. aeruginosa</i> PAO1		NC_002516.2	Feb. 2015
P126	<i>P. aeruginosa</i> PAO1H2O		CP008749.1	Feb. 2015
P127	<i>P. aeruginosa</i> PAO1-VE13		NC_022594.1	Feb. 2015
P128	<i>P. aeruginosa</i> PAO1-VE2		NC_022591.1	Feb. 2015
P129	<i>P. aeruginosa</i> PAO581		NC_022361.1	Feb. 2015
P130	<i>P. aeruginosa</i> PSE305		HG974234.1	Feb. 2015
P131	<i>P. aeruginosa</i> RP73		NC_021577.1	Feb. 2015
P132	<i>P. aeruginosa</i> SCV20265		NC_023149.1	Feb. 2015
P133	<i>P. aeruginosa</i> UCBPP-PA14		NC_008463.1	Feb. 2015
P134	<i>P. aeruginosa</i> YL84		CP007147.1	Feb. 2015
P135	<i>P. aestusnigri</i> CECT 8317 (T)	FNVE01		Apr. 2017
P136	<i>P. agarici</i> NCPPB 2289 (T)	AKBQ01 WGS_SCAFLD		Feb. 2015
P137	<i>P. alcaligenes</i> MRY13-0052	BATO01		Feb. 2015
P138	<i>P. alcaligenes</i> NBRC 14159 (T)	BATIO1		Feb. 2015

ID	Stammbezeichner <sub>NCBI</sub>	WGS Kennung	RefSeq/GenBank Sequenzen	Datum Bezug
P139	<i>P. alcaliphila</i> 34	ANGB01		Feb. 2015
P140	<i>P. alcaliphila</i> JCM 10630 (T)	FNAE01		Apr. 2017
P141	<i>P. alkylphenolica</i> KL28 (T)		NZ_CP009048.1	Apr. 2017
P142	<i>P. amygdali</i> ICMP 3918 (T)	LJPQ01		Apr. 2017
P143	<i>P. anguilliseptica</i> DSM 12111 (T)	FNSC01		Apr. 2017
P144	<i>P. argentinensis</i> LMG 22563 (T)	FORC01		Apr. 2017
P145	<i>P. arsenicoxydans</i> CECT 7543 (T)		LT629705.1	Apr. 2017
P146	<i>P. asplenii</i> ATCC 23835 (T)		LT629777.1	Apr. 2017
P147	<i>P. asturiensis</i> LMG 26898 (T)	FRDA01		Apr. 2017
P148	<i>P. avellanae</i> BPIC 631 (T)	AKBS01	WGS_SCAFLD	Feb. 2015
P149	<i>P. azotifigens</i> DSM 17556 (T)	AUDU01		Feb. 2015
P150	<i>P. azotoformans</i> LMG 21611 (T)		NZ_LT629702.1	Apr. 2017
P151	<i>P. balearica</i> DSM 6083 (T)		CP007511.1	Feb. 2015
P152	<i>P. bauzanensis</i> DSM 22558 (T)	FOGN01		Apr. 2017
P153	<i>P. bauzanensis</i> W13Z2	JFHS01		Feb. 2015
P154	<i>P. benzenivorans</i> DSM 8628 (T)	FNCT01		Apr. 2017
P155	<i>P. borbori</i> DSM 17834 (T)	FOWX01		Apr. 2017
P156	<i>P. brassicacearum</i> DF41		CP007410.1	Feb. 2015
P157	<i>P. brassicacearum</i> NFM421		NC_015379.1	Feb. 2015
P158	<i>P. brassicacearum</i> PA1G7	JBON01		Feb. 2015
P159	<i>P. brassicacearum</i> PP1_210F	AYJR01		Feb. 2015
P160	<i>P. caeni</i> DSM 24390 (T)	ATXQ01		Feb. 2015
P161	<i>P. "canadensis"</i> 2-92 (iT)	AYTD01		Apr. 2017
P162	<i>P. cannabina</i> ICMP 2823 (T)	FNKU01		Apr. 2017
P163	<i>P. caricapapayae</i> ICMP 2855 (T)	LJPW01		Apr. 2017
P164	<i>P. cedrina</i> (ssp. <i>cedrina</i> ) DSM 17516 (T)	MNPW01		Apr. 2017
P165	<i>P. "cerasi"</i> 58 (iT)		NZ_LT222313.1 bis NZ_LT222319.1	Apr. 2017
P166	<i>P. chloritidismutans</i> AW-1	AOFQ01		Feb. 2015
P167	<i>P. chlororaphis</i> (ssp. <i>aurantiaca</i> ) LMG 21630 (T)		NZ_LT629747.1	Apr. 2017
P168	<i>P. chlororaphis</i> (ssp. <i>aureofaciens</i> ) NBRC 3521 (T)	BBQB01		Feb. 2015
P169	<i>P. chlororaphis</i> (ssp. <i>piscium</i> ) DSM 21509 (T)	LHUZ01		Apr. 2017
P170	<i>P. chlororaphis</i> 30-84	AHHJ01		Feb. 2015
P171	<i>P. chlororaphis</i> EA105	JSFK01		Feb. 2015
P172	<i>P. chlororaphis</i> HT66	ATBG01		Feb. 2015
P173	<i>P. chlororaphis</i> JD37		CP009290.1	Feb. 2015
P174	<i>P. chlororaphis</i> O6	AHOT01		Feb. 2015
P175	<i>P. chlororaphis</i> PA23		CP008696.1	Feb. 2015
P176	<i>P. chlororaphis</i> PB-St2	AYUD01		Feb. 2015
P177	<i>P. chlororaphis</i> YL-1	AWWJ01		Feb. 2015
P178	<i>P. cichorii</i> ATCC 10857 (T)	FNIK01		Apr. 2017

ID	Stammbezeichner <sub>NCBI</sub>	WGS Kennung	RefSeq/GenBank Sequenzen	Datum Bezug
P179	<i>P. cichorii</i> JBC1		CP007039.1	Feb. 2015
P180	<i>P. citronellolis</i> LMG 18378 (T)	FOLS01		Apr. 2017
P181	<i>P. coleopterorum</i> LMG 28558 (T)	FNTZ01		Apr. 2017
P182	<i>P. composti</i> CCUG 59231 (T)	FOWP01		Apr. 2017
P183	<i>P. congelans</i> DSM 14939 (T)	FNJH01		Apr. 2017
P184	<i>P. corrugata</i> CFBP 5454	ATKI01		Feb. 2015
P185	<i>P. corrugata</i> DSM 7228 (T)	LHVK01		Apr. 2017
P186	<i>P. constantinii</i> LMG 22119 (T)	MDDR01		Apr. 2017
P187	<i>P. cremoricolorata</i> DSM 17059 (T)	AUEA01		Feb. 2015
P188	<i>P. cremoricolorata</i> ND07		CP009455.1	Feb. 2015
P189	<i>P. cuatrocienegasensis</i> CIP 109853 (T)	FOFP01		Apr. 2017
P190	<i>P. delhiensis</i> CCM 7361 (T)	FNEC01		Apr. 2017
P191	<i>P. donghuensis</i> HYS (T)	AJJP01		Apr. 2017
P192	<i>P. endophytica</i> BSTT44 (T)	LLWH01		Apr. 2017
P193	<i>P. entomophila</i> L48 (T)		NC_008027.1	Feb. 2015
P194	<i>P. extremorientalis</i> LMG 19695 (T)	MDGK01		Apr. 2017
P195	<i>P. ficuserectae</i> ICMP 7848 (T)	LJQJ01		Apr. 2017
P196	<i>P. flavescens</i> NBRC 103044 (T)	BCZZ01		Apr. 2017
P197	<i>P. flexibilis</i> CGMCC 1.1365 (T)	FMUP01		Apr. 2017
P198	<i>P. fluorescens</i> 2-79	JXCQ01		Feb. 2015
P199	<i>P. fluorescens</i> A506		NC_017911.1, NC_021361.1	Feb. 2015
P200	<i>P. fluorescens</i> ATCC 17400	JENC01		Feb. 2015
P201	<i>P. fluorescens</i> AU14440	JRXX01		Feb. 2015
P202	<i>P. fluorescens</i> AU14705	JRYA01		Feb. 2015
P203	<i>P. fluorescens</i> AU2989	JRXT01		Feb. 2015
P204	<i>P. fluorescens</i> BBc6R8	AKXH02		Feb. 2015
P205	<i>P. fluorescens</i> BRIP34879	AMZW01		Feb. 2015
P206	<i>P. fluorescens</i> BS2	AMZG01		Feb. 2015
P207	<i>P. fluorescens</i> C2	JSAK01		Feb. 2015
P208	<i>P. fluorescens</i> EGD-AQ6	AVQG01		Feb. 2015
P209	<i>P. fluorescens</i> F113		NC_016830.1	Feb. 2015
P210	<i>P. fluorescens</i> FH5	AOJA01		Feb. 2015
P211	<i>P. fluorescens</i> H16	JSAL01		Feb. 2015
P212	<i>P. fluorescens</i> HK44	AFOY02		Feb. 2015
P213	<i>P. fluorescens</i> LMG 5329	ASGY01		Feb. 2015
P214	<i>P. fluorescens</i> MEP34	JXQY01		Feb. 2015
P215	<i>P. fluorescens</i> NCIMB 11764		NZ_CM001560.1	Feb. 2015
P216	<i>P. fluorescens</i> PA3G8	JBOO01		Feb. 2015
P217	<i>P. fluorescens</i> PA4C2	AXDA01		Feb. 2015
P218	<i>P. fluorescens</i> PCL1751		NZ_CP010896.1	Feb. 2015

ID	Stammbezeichnung <sub>NCBI</sub>	WGS Kennung	RefSeq/GenBank Sequenzen	Datum Bezug
P219	<i>P. fluorescens</i> Pf0-1		NC_007492.2	Feb. 2015
P220	<i>P. fluorescens</i> Pf29Arp	ANOR01		Feb. 2015
P221	<i>P. fluorescens</i> Q2-87		NZ_CM001558.1	Feb. 2015
P222	<i>P. fluorescens</i> Q8r1-96		NZ_CM001512.1	Feb. 2015
P223	<i>P. fluorescens</i> R124		NZ_CM001561.1, NZ_CM001562.1	Feb. 2015
P224	<i>P. fluorescens</i> SBW25		NC_012660.1	Feb. 2015
P225	<i>P. fluorescens</i> SF39a	JTGG01		Feb. 2015
P226	<i>P. fluorescens</i> SF4c	JTGH01		Feb. 2015
P227	<i>P. fluorescens</i> SRM1	CDMF01		Feb. 2015
P228	<i>P. fluorescens</i> SS101		NZ_CM001513.1	Feb. 2015
P229	<i>P. fluorescens</i> UK4		CP008896.1	Feb. 2015
P230	<i>P. fluorescens</i> WH6		NZ_CM001025.1	Feb. 2015
P231	<i>P. formosensis</i> JCM 18415 (T)	FOYD01		Apr. 2017
P232	<i>P. fragi</i> A22	AHZY01		Feb. 2015
P233	<i>P. fragi</i> B25 (T)	AHZX01		Feb. 2015
P234	<i>P. frederiksbergensis</i> S18	JQGJ01		Feb. 2015
P235	<i>P. fulva</i> 12-X		NC_015556.1	Feb. 2015
P236	<i>P. fulva</i> MEJ086	JXQW01		Feb. 2015
P237	<i>P. fulva</i> NBRC 16637 (T)	JHYU01		Feb. 2015
P238	<i>P. fuscovaginae</i> CB98818	ALAQ01		Feb. 2015
P239	<i>P. fuscovaginae</i> LMG 2158 (T)		NZ_LT629972.1	Apr. 2017
P240	<i>P. fuscovaginae</i> UPB0736	AIEU01		Feb. 2015
P241	<i>P. graminis</i> DSM 11363 (T)	FOHW01		Apr. 2017
P242	<i>P. granadensis</i> LMG 27940 (T)		LT629778.1	Apr. 2017
P243	<i>P. guangdongensis</i> CCTCC 2012022 (T)		LT629780.1	Apr. 2017
P244	<i>P. guariconensis</i> LMG 27394 (T)	FMYX01		Apr. 2017
P245	<i>P. guguanensis</i> JCM 18416 (T)	FNJJ01		Apr. 2017
P246	<i>P. guineae</i> LMG 24016 (T)	FOQL01		Apr. 2017
P247	<i>P. hussainii</i> JCM 19513 (T)	FOAS01		Apr. 2017
P248	<i>P. indica</i> NBRC 103045 (T)	BDAC01		Apr. 2017
P249	<i>P. japonica</i> NBRC 103040 (T)	BBIR01		Feb. 2015
P250	<i>P. jinjuensis</i> NBRC 103047 (T)	BDAD01		Apr. 2017
P251	<i>P. kilonensis</i> DSM 13647 (T)	LHVH01		Apr. 2017
P252	<i>P. knackmussii</i> B13 (T)		HG322950.1	Feb. 2015
P253	<i>P. kunmingensis</i> DSM 25974 (T)	FORS01		Apr. 2017
P254	<i>P. kuykendallii</i> NRRL B-59562 (T)	FNNU01		Apr. 2017
P255	<i>P. linyingensis</i> LMG 25967 (T)	FNZE01		Apr. 2017
P256	<i>P. litoralis</i> 2SM5 (T)		LT629748.1	Apr. 2017
P257	<i>P. lutea</i> DSM 17257 (T)	JRMB01		Feb. 2015
P258	<i>P. luteola</i> NBRC 103146 (T)	BDAE01		Apr. 2017

ID	Stammbezeichner <sub>NCBI</sub>	WGS Kennung	RefSeq/GenBank Sequenzen	Datum Bezug
P259	<i>P. luteola</i> XLDN4-9	ALAT01		Feb. 2015
P260	<i>P. mandelii</i> JR-1		CP005960.1, CP005961.1	Feb. 2015
P261	<i>P. marincola</i> JCM 14761 (T)	FPBC01		Apr. 2017
P262	<i>P. mediterranea</i> CFBP 5447 (T)	AUPB01		Feb. 2015
P263	<i>P. meliae</i> CFBP 3225 (T)	JYHE01		Apr. 2017
P264	<i>P. mendocina</i> DLHK	ALKM01		Feb. 2015
P265	<i>P. mendocina</i> EGD-AQ5	AVQF01		Feb. 2015
P266	<i>P. mendocina</i> NBRC 14162 (T)	BBQC01		Feb. 2015
P267	<i>P. mendocina</i> NK-01		NC_015410.1	Feb. 2015
P268	<i>P. mendocina</i> S13.2	JOVT01		Feb. 2015
P269	<i>P. mendocina</i> S5.2	AMCD02		Feb. 2015
P270	<i>P. mendocina</i> ymp		NC_009439.1	Feb. 2015
P271	<i>P. migulae</i> NBRC 103157 (T)	BDAG01		Apr. 2017
P272	<i>P. mohnii</i> DSM 18327 (T)	FNRV01		Apr. 2017
P273	<i>P. montellii</i> NBRC 103158 (T)	JHYV01		Feb. 2015
P274	<i>P. montellii</i> SB3078		NC_023075.1	Feb. 2015
P275	<i>P. montellii</i> SB3101		NC_023076.1	Feb. 2015
P276	<i>P. moraviensis</i> R28-S	AYMZ01		Feb. 2015
P277	<i>P. mosselii</i> DSM 17497 (T)	JHYW01		Feb. 2015
P278	<i>P. mosselii</i> SJ10		CP009365.1	Feb. 2015
P279	<i>P. mucidolens</i> NBRC 103159 (T)	BDAH01		Apr. 2017
P280	<i>P. nitroreducens</i> Aramco J	JUEH01		Feb. 2015
P281	<i>P. nitroreducens</i> HBP1	AZRU01		Feb. 2015
P282	<i>P. oleovorans</i> (ssp. <i>lubricantis</i> ) RS1 (T)	FNZC01		Apr. 2017
P283	<i>P. oleovorans</i> (ssp. <i>oleovorans</i> ) NBRC 13583 (T)	BDAL01		Apr. 2017
P284	<i>P. oleovorans</i> MGY01	JNHE01		Feb. 2015
P285	<i>P. oleovorans</i> MOIL14HWK12	AZOB01		Feb. 2015
P286	<i>P. "oryzae"</i> KCTC 32247 (iT)		LT629751.1	Apr. 2017
P287	<i>P. otitidis</i> DSM 17224 (T)	FOJP01		Apr. 2017
P288	<i>P. otitidis</i> LNU-E-001	JGYF01		Feb. 2015
P289	<i>P. pachastrellae</i> CCUG 46540 (T)	MUBC01		Apr. 2017
P290	<i>P. panipatensis</i> CCM 7469 (T)	FNDS01		Apr. 2017
P291	<i>P. parafulva</i> CRS01-1		CP009747.1	Feb. 2015
P292	<i>P. parafulva</i> DSM 17004 (T)	AUEB01		Feb. 2015
P293	<i>P. pelagia</i> CL-AP6 (T)	AROI01		Feb. 2015
P294	<i>P. perfectomarina</i> ATCC 14405	AGSL01		Feb. 2015
P295	<i>P. plecoglossida</i> NB2011	ASJX01		Feb. 2015
P296	<i>P. plecoglossida</i> NBRC 103162 (T)	JHYX01		Feb. 2015
P297	<i>P. plecoglossida</i> NyZ12		NZ_CP010359.1	Feb. 2015
P298	<i>P. plecoglossida</i> TND35	JOJY01		Feb. 2015

ID	Stammbezeichner <sub>NCBI</sub>	WGS Kennung	RefSeq/GenBank Sequenzen	Datum Bezug
P299	<i>P. poae</i> RE*1-1-14		NC_020209.1	Feb. 2015
P300	<i>P. pohangensis</i> DSM 17875 (T)		LT629785.1	Apr. 2017
P301	<i>P. prosekii</i> LMG 26867 (T)		LT629762.1	Apr. 2017
P302	<i>P. protegens</i> Cab57		NZ_AP014522.1	Feb. 2015
P303	<i>P. protegens</i> CHA0 (T)		NC_021237.1	Feb. 2015
P304	<i>P. protegens</i> Pf-5		NC_004129.6	Feb. 2015
P305	<i>P. pseudoalcaligenes</i> AD6	JFJN01		Feb. 2015
P306	<i>P. pseudoalcaligenes</i> CECT 5344	CAIG01		Feb. 2015
P307	<i>P. pseudoalcaligenes</i> KF707	AJMR01		Feb. 2015
P308	<i>P. pseudoalcaligenes</i> NBRC 14167	BDAJ01		Apr. 2017
P309	<i>P. psychrophila</i> HA-4	ALJC01		Feb. 2015
P310	<i>P. psychrotolerans</i> L19	AHBD01		Feb. 2015
P311	<i>P. punonensis</i> CECT 8089 (T)	FRBQ01		Apr. 2017
P312	<i>P. putida</i> B6-2	AGCS01		Feb. 2015
P313	<i>P. putida</i> BIRD-1		NC_017530.1	Feb. 2015
P314	<i>P. putida</i> CSV86	AMWJ01		Feb. 2015
P315	<i>P. putida</i> DLL-E4		CP007620.1	Feb. 2015
P316	<i>P. putida</i> F1		NC_009512.1	Feb. 2015
P317	<i>P. putida</i> GB-1		NC_010322.1	Feb. 2015
P318	<i>P. putida</i> H8234		NC_021491.1	Feb. 2015
P319	<i>P. putida</i> HB3267		NC_019905.1, NC_019906.1	Feb. 2015
P320	<i>P. putida</i> KF703	BBQL01		Feb. 2015
P321	<i>P. putida</i> KT2440		NC_002947.3	Feb. 2015
P322	<i>P. putida</i> LS46	ALPV02		Feb. 2015
P323	<i>P. putida</i> MTCC5279	AMZE01		Feb. 2015
P324	<i>P. putida</i> NBRC 14164 (T)		NC_021505.1	Feb. 2015
P325	<i>P. putida</i> ND6		NC_017986.1, NC_018746.1	Feb. 2015
P326	<i>P. putida</i> OUS82	AZBL01		Feb. 2015
P327	<i>P. putida</i> PA14H7	JBOP01		Feb. 2015
P328	<i>P. putida</i> PD1	JUHC01		Feb. 2015
P329	<i>P. putida</i> S11	ALXA01		Feb. 2015
P330	<i>P. putida</i> S12		NZ_CP009974.1, NZ_CP009975.1	Feb. 2015
P331	<i>P. putida</i> S13.1.2	AVFB02		Feb. 2015
P332	<i>P. putida</i> S16		NC_015733.1	Feb. 2015
P333	<i>P. putida</i> S610	AYJQ01		Feb. 2015
P334	<i>P. putida</i> SJTE-1	AKCL01		Feb. 2015
P335	<i>P. putida</i> TRO1	APBQ01		Feb. 2015
P336	<i>P. putida</i> UASWS0946	JXOG01		Feb. 2015
P337	<i>P. putida</i> W15Oct28	JENB01		Feb. 2015
P338	<i>P. putida</i> W619		NC_010501.1	Feb. 2015

ID	Stammbezeichner <sub>NCBI</sub>	WGS Kennung	RefSeq/GenBank Sequenzen	Datum Bezug
P339	<i>P. putida</i> YKD221	BBNC01		Feb. 2015
P340	<i>P. reinekei</i> MT-1 (T)	MSTQ01		Apr. 2017
P341	<i>P. resinovorans</i> DSM 21078 (T)	AUIE01		Feb. 2015
P342	<i>P. resinovorans</i> NBRC 106553		NC_021499.1, NC_021506.1	Feb. 2015
P343	<i>P. rhizosphaerae</i> DSM 16299 (T)		CP009533.1	Feb. 2015
P344	<i>P. rhodesiae</i> FF9	CCYI01		Feb. 2015
P345	<i>P. sabulinigri</i> JCM 14963 (T)		LT629763.1	Apr. 2017
P346	<i>P. sagittaria</i> JCM 18195 (T)	FOXMO1		Apr. 2017
P347	<i>P. salegens</i> CECT 8338 (T)		LT629787.1	Apr. 2017
P348	<i>P. salomonii</i> ICMP 14252 (T)	FNOX01		Apr. 2017
P349	<i>P. saponiphila</i> DSM 9751 (T)	FNTJ01		Apr. 2017
P350	<i>P. savastanoi</i> ICMP 4352 (T)	LJRJ01		Apr. 2017
P351	<i>P. seleniipraecipitans</i> LMG 25475 (T)	FNBM01		Apr. 2017
P352	<i>P. "sihuiensis"</i> KCTC 32246 (iT)		LT629797.1	Apr. 2017
P353	<i>P. simiae</i> CCUG 50988 (T)	FOKB01		Apr. 2017
P354	<i>P. simiae</i> 2-36	JRMC01		Feb. 2015
P355	<i>P. simiae</i> MEB105	JXQT01		Feb. 2015
P356	<i>P. simiae</i> WCS417		NZ_CP007637.1	Feb. 2015
P357	<i>P. soli</i> LMG 27941 (T)	FOEQ01		Apr. 2017
P358	<i>P. sp. (batumici)</i> UCM B-321	JXDG01		Apr. 2017
P359	<i>P. sp. (capeferrum)</i> WCS358	JMIT01		Apr. 2017
P360	<i>P. sp. (defensor)</i> WCS374		NZ_CP007638.1	Apr. 2017
P361	<i>P. sp. (saudiphocaensis)</i> 20_BN	CCSF01		Apr. 2017
P362	<i>P. sp.</i> BICA1-14	JUEC01		Feb. 2015
P363	<i>P. sp.</i> TX1	AMZB01		Feb. 2015
P364	<i>P. straminea</i> JCM 2783 (T)	FOMO01		Apr. 2017
P365	<i>P. stutzeri</i> 19SMN4		CP007509.1, CP007510.1	Feb. 2015
P366	<i>P. stutzeri</i> 28a24		CP007441.1	Feb. 2015
P367	<i>P. stutzeri</i> A1501		NC_009434.1	Feb. 2015
P368	<i>P. stutzeri</i> ATCC 17588 (T)		NC_015740.1	Feb. 2015
P369	<i>P. stutzeri</i> B1SMN1	AMVM01		Feb. 2015
P370	<i>P. stutzeri</i> CCUG 29243		NC_018028.1	Feb. 2015
P371	<i>P. stutzeri</i> DSM 10701		NC_018177.1	Feb. 2015
P372	<i>P. stutzeri</i> DSM 4166		NC_017532.1	Feb. 2015
P373	<i>P. stutzeri</i> KOS6	AMCZ02		Feb. 2015
P374	<i>P. stutzeri</i> NF13	AOBS01		Feb. 2015
P375	<i>P. stutzeri</i> RCH2		NC_019936.1 bis NC_019939.1	Feb. 2015
P376	<i>P. stutzeri</i> T13	ALJB01		Feb. 2015
P377	<i>P. stutzeri</i> TS44	AJXE01		Feb. 2015
P378	<i>P. stutzeri</i> XLDN-R	AKYE01		Feb. 2015

ID	Stammbezeichner <sub>NCBI</sub>	WGS Kennung	RefSeq/GenBank Sequenzen	Datum Bezug
P379	<i>P. synxantha</i> BG33R		NZ_CM001514.1	Feb. 2015
P380	<i>P. syringae</i> (ssp. <i>syringae</i> ) DSM 10604 (T)	JALK01		Feb. 2015
P381	<i>P. syringae</i> 1212	AVCR02		Feb. 2015
P382	<i>P. syringae</i> 1448A		NC_005773.3, NC_007274.1, NC_007275.1	Feb. 2015
P383	<i>P. syringae</i> 642	ADGB01		Feb. 2015
P384	<i>P. syringae</i> 6605	AJXI01		Feb. 2015
P385	<i>P. syringae</i> Alf3	JPNN01		Feb. 2015
P386	<i>P. syringae</i> B076	AEGG01		Feb. 2015
P387	<i>P. syringae</i> B301D-R	JALJ01		Feb. 2015
P388	<i>P. syringae</i> B576	JRUO01		Feb. 2015
P389	<i>P. syringae</i> B64		NZ_CM001763.1	Feb. 2015
P390	<i>P. syringae</i> B728a		NC_007005.1	Feb. 2015
P391	<i>P. syringae</i> BRIP34876	AMXK01		Feb. 2015
P392	<i>P. syringae</i> BRIP34881	AMXL01		Feb. 2015
P393	<i>P. syringae</i> BRIP39023	AMZX01		Feb. 2015
P394	<i>P. syringae</i> CC1417	AVEO02		Feb. 2015
P395	<i>P. syringae</i> CC1466	AVEM02		Feb. 2015
P396	<i>P. syringae</i> CC1513	AVEL02		Feb. 2015
P397	<i>P. syringae</i> CC1524	AVEK02		Feb. 2015
P398	<i>P. syringae</i> CC1557		CP007014.1, CP007015.1	Feb. 2015
P399	<i>P. syringae</i> CC1583	AVEF02		Feb. 2015
P400	<i>P. syringae</i> CC1629	AVEE02		Feb. 2015
P401	<i>P. syringae</i> CC1630	AVED02		Feb. 2015
P402	<i>P. syringae</i> CC440	AVEC02		Feb. 2015
P403	<i>P. syringae</i> CEB003	JPQT01		Feb. 2015
P404	<i>P. syringae</i> GAW0119	JPQU01		Feb. 2015
P405	<i>P. syringae</i> ICMP 18804	ANJE01		Feb. 2015
P406	<i>P. syringae</i> ICMP 18806	ANJF01		Feb. 2015
P407	<i>P. syringae</i> ICMP 18807	ANJL01		Feb. 2015
P408	<i>P. syringae</i> ICMP 18884	AOKO01		Feb. 2015
P409	<i>P. syringae</i> ICMP 19072	AOJW01		Feb. 2015
P410	<i>P. syringae</i> ICMP 19098	AOKE01		Feb. 2015
P411	<i>P. syringae</i> ICMP 19099	AOKD01		Feb. 2015
P412	<i>P. syringae</i> ICMP 9617	AOKP01		Feb. 2015
P413	<i>P. syringae</i> ISPaVe013	AKCJ01		Feb. 2015
P414	<i>P. syringae</i> ISPaVe037	AKCK01		Feb. 2015
P415	<i>P. syringae</i> KCTC 12500	AYTM01		Feb. 2015
P416	<i>P. syringae</i> LMG 2367	ALAC01		Feb. 2015
P417	<i>P. syringae</i> PP1	AUZR02		Feb. 2015
P418	<i>P. syringae</i> race 4	AEGH01		Feb. 2015

<b>ID</b>	<b>Stammbezeichner<sub>NCBI</sub></b>	<b>WGS Kennung</b>	<b>RefSeq/GenBank Sequenzen</b>	<b>Datum Bezug</b>
P419	<i>P. syringae</i> SM		NZ_CM001986.1, NZ_CM001987.1	Feb. 2015
P420	<i>P. syringae</i> UB0390	JPQV01		Feb. 2015
P421	<i>P. syringae</i> USA011	AVDX02		Feb. 2015
P422	<i>P. syringae</i> yuexi-1	JWJF01		Feb. 2015
P423	<i>P. taeaanensis</i> MS-3 (T)	AWSQ01		Feb. 2015
P424	<i>P. taiwanensis</i> DSM 21245 (T)	AUEC01		Feb. 2015
P425	<i>P. thermotolerans</i> DSM 14292 (T)	AQPA01		Feb. 2015
P426	<i>P. thermotolerans</i> J53	AZUT01		Feb. 2015
P427	<i>P. thivervalensis</i> DSM 13194 (T)	LHVE01		Apr. 2017
P428	<i>P. tolaasii</i> NCPPB 2192 (T)		NZ_CP020369.1	Apr. 2017
P429	<i>P. toyotomiensis</i> JCM 15604 (T)	FOXK01		Apr. 2017
P430	<i>P. tremae</i> ICMP 9151 (T)	LJRO01		Apr. 2017
P431	<i>P. tuomuerensis</i> JCM 14085 (T)	JTAK01		Feb. 2015
P432	<i>P. veronii</i> 1YdBTEX2	AOUH01		Feb. 2015
P433	<i>P. "versuta"</i> L10.10 (iT)		NZ_CP012676.1	Apr. 2017
P434	<i>P. viridiflava</i> DSM 6694 (T)	JRXH01		Apr. 2017
P435	<i>P. viridiflava</i> LMCA8	JXQO01		Feb. 2015
P436	<i>P. viridiflava</i> UASWS0038	AMQP01		Feb. 2015
P437	<i>P. vranovensis</i> DSM 16006 (T)	AUED01		Feb. 2015
P438	<i>P. xanthomarina</i> DSM 18231 (T)	FQXA01		Apr. 2017
P439	<i>P. xanthomarina</i> S11	CCYE01		Feb. 2015
P440	<i>P. yamanorum</i> LMG 27247 (T)		LT629793.1	Apr. 2017
P441	<i>P. zeshuii</i> KACC 15471 (T)	FQYS01		Apr. 2017

**ID**, Kennung zur eindeutigen Identifizierung eines Stammes

**Stammbezeichner<sub>NCBI</sub>**, Bezeichnung des Stammes, wobei Gattungs- und Artzugehörigkeit sowie die Stammkennung vom NCBI übernommen wurden

**WGS Kennung**, Kennung des WGS Projektes bestehend aus einem 4 buchstabigen Präfix (Projekt) gefolgt von 2 Ziffern (Versionsnummer)

**RefSeq/Genbank Sequenzen**, Kennungen der NCBI oder Genbank Referenzgenomsequenzen

**Datum Bezug**, Zeitpunkt des Bezuges der Genomsequenzen vom NCBI

(T), anerkannter Typstamm nach LPSN [67] (Stand Apr. 2017)

(iT), nicht anerkannter Typstamm, da Artnamen (noch) nicht in LPSN [67] aufgenommen (Stand Apr. 2017)

WGS\_SCAFLD, Bezug der WGS Scaffold Sequenzen

Tabelle 29 Verworfenne *Pseudomonas* NCBI Genomdaten

Stammbezeichner <sub>NCBI</sub>	WGS	RefSeq/GenBank Sequenzen	Grund des Ausschlusses
<i>P. alcaligenes</i> OT 69	ATCP01		keine 16S
<i>P. brassicacearum</i> 51MFCVI2.1	AZOC01		16S zu kurz
<i>P. fluorescens</i> AU10973	JRXV01		16S zu kurz
<i>P. fluorescens</i> AU11518	JRXW01		16S zu kurz
<i>P. fluorescens</i> AU12271	JRYP01		16S zu kurz
<i>P. fluorescens</i> AU14917	JRXY01		16S zu kurz
<i>P. fluorescens</i> AU6026	JRXU01		16S zu kurz
<i>P. fluorescens</i> AU6308	JRXZ01		16S zu kurz
<i>P. fluorescens</i> DSM 8569	JXOE01		keine 16S
<i>P. fluorescens</i> Wayne1	CADX01		16S zu kurz
<i>P. geniculata</i> N1	AJLO01		kein <i>Pseudomonas</i>
<i>P. mandelii</i> 36MFCvi1.1	ARLP01		16S zu kurz
<i>P. mandelii</i> PD30	AZQQ01		16S zu kurz
<i>P. monteilii</i> QM	AHGZ01		keine 16S
<i>P. putida</i> B001	CAED01		16S zu kurz
<i>P. putida</i> DOT-T1E		NC_018220.1	untypische 16S
<i>P. putida</i> LF54	AOUR02		16S zu kurz
<i>P. stutzeri</i> MF28	ATAR01		16S zu kurz
<i>P. stutzeri</i> SDM-LAC	AGSX01		16S zu kurz
<i>P. stutzeri</i> YC-YH1	JUDR01		kein <i>rpoD</i> -Gen
<i>P. umsongensis</i> 20MFCvi1.1	ARIW01		16S zu kurz
<i>P. umsongensis</i> UNC430CL58Col	JHVT01		16S zu kurz

**Stammbezeichner<sub>NCBI</sub>**, Bezeichnung des Stammes, wobei Gattungs- und Artzugehörigkeit sowie die Stammkennung vom NCBI übernommen wurden

**WGS Kennung**, Kennung des WGS Projektes bestehend aus einem 4 buchstabigen Präfix (Projekt) gefolgt von 2 Ziffern (Versionsnummer)

**RefSeq/Genbank Sequenzen**, Kennungen der NCBI oder Genbank Referenzgenomsequenzen

**Grund des Ausschlusses**, Grund weshalb die NCBI Genomassemblierungen keine weitere Verwendung fanden

Tabelle 30 Verwendete *Pseudomonas* NCBI Einzelgensequenzen

ID	Stammbezeichner <sub>NCBI</sub>	16S	<i>gyrB</i>	<i>rpoB</i>	<i>rpoD</i>
		Sequenz Stammkennung	Sequenz Stammkennung	Sequenz Stammkennung	Sequenz Stammkennung
P442	<i>P. antarctica</i> LMG 22709 (T)	AJ537601.1   CMS 35 (T)	FN554169.1   LMG 22709 (T)	FN554727.1   LMG 22709 (T)	FN554450.1   LMG 22709 (T)
P443	<i>P. baetica</i> CECT 7720 (T)	FM201274.3   a390 (T)	HE800470.1   CECT 7720 (T)	HE800504.1   CECT 7720 (T)	FN678357.1   a390 (T)
P444	<i>P. brassicacearum</i> (ssp. <i>brassicacearum</i> ) CFBP 11706 (T)	NR_024950.1   DBK11 (T)	AM084675.1   CFBP 11706 (T)	AJ717436.1   CIP 107059 (T)	AM084334.1   CFBP 11706 (T)
P445	<i>P. cedrina</i> (ssp. <i>fulgida</i> ) DSM 14938 (T)	KX186939.1   DSM 14938 (T)	KP260626.1   DSM 14938 (T)	KX186859.1   DSM 14938 (T)	HE586449.1   LMG 21467(T)
P446	<i>P. duriflava</i> KCTC 22129 (T)	EU046271.1   HR2 (T)	HE800478.1   KCTC 22129 (T)	HE800512.1   KCTC 22129 (T)	HE800494.1   KCTC 22129 (T)
P447	<i>P. frederiksbergensis</i> DSM 13022 (T)	AJ249382.1   JAJ28 (T)	AM084676.1   DSM 13022 (T)	AJ717465.1   CIP 106887 (T)	AM084335.2   DSM 13022 (T)
P448	<i>P. helmanticensis</i> OHA11 (T)	NR_126220.1   OHA11 (T)	LT601023.1   CECT 8548 (T)	HG940518.1   OHA11 (T)	HG940517.1   OHA11 (T)
P449	<i>P. jessenii</i> CIP 105274 (T)	AF068259.1   CIP 105274 (T)	FN554191.1   CIP 105274 (T)	AJ717447.1   CIP 105274 (T)	FN554473.1   CIP 105274 (T)
P450	<i>P. lurida</i> P513-18 (T)	NR_042199.1   P513-18 (T)	HE800480.1   P513-18 (T)	HE800513.1   P513-18 (T)	HE800497.1   P513-18 (T)
P451	<i>P. meridiana</i> CIP 108465 (T)	AJ537602.1   CMS 38 (T)	FN554203.1   CIP 108465 (T)	FN554740.1   CIP 108465 (T)	FN554485.1   CIP 108465 (T)
P452	<i>P. moorei</i> CCUG 53114 (T)	AM293566.1   RW10 (T)	AM293560.1   RW10 (T)	FN554742.1   CCUG 53114 (T)	FN554489.1   CCUG 53114 (T)
P453	<i>P. moraviensis</i> DSM 16007 (T)	AY970952.1   CCM 7280 (T)	FN554206.1   DSM 16007 (T)	FN554743.1   DSM 16007 (T)	FN554490.1   DSM 16007 (T)
P454	<i>P. palleroniana</i> LMG 23076 (T)	AY091527.1   CFBP 4389 (T)	FN554213.1   LMG 23076 (T)	FN554747.1   LMG 23076 (T)	FN554497.1   LMG 23076 (T)
P455	<i>P. pertucinogena</i> LMG 1874 (T)	AB021380.1   IFO 14163 (T)	DQ350613.2   JCM 11950 (T)	AJ717441.1   LMG 1874 (T)	FN554502.1   LMG 1874 (T)
P456	<i>P. salina</i> XCD-X85 (T)	KC762324   XCD-X85 (T)	KP098570   XCD-X85 (T)	KP075008   XCD-X85 (T)	KP075009   XCD-X85 (T)
P457	<i>P. segetis</i> IMSNU 14100 (T)	AY770691.2   FR1439 (T)	HE800484.1   IMSNU 14101 (T)	HE800516.1   IMSNU 14101 (T)	HE800500.1   IMSNU 14101 (T)
P458	<i>P. turukhanskensis</i> IB1.1 (T)	KP306892.1   IB1.1 (T)	LT219440.1   IB1.1 (T)	LT219439.1   IB1.1 (T)	LT219438.1   IB1.1 (T)
P459	<i>P. umsongensis</i> LMG 21317 (T)	NR_025227.1   Ps 3-10 (T)	FN554231.1   LMG 21317 (T)	FN554763.1   LMG 21317 (T)	FN554516.1   LMG 21317 (T)
P460	<i>P. vancouverensis</i> ATCC 700688 (T)	NR_041953.1   DhA-51 (T)	FN554232.1   ATCC 700688 (T)	AJ717473.1   CIP 106707 (T)	FN554517.1   ATCC 700688 (T)
P136	<i>P. agarici</i> NCPPB 2289 (T)	Z76652.1   LMG 2112 (T)	---	---	---
P149	<i>P. azotifigens</i> DSM 17556 (T)	AB189452.1   6H33b (T)	---	---	---
P155	<i>P. borbori</i> DSM 17834 (T)	AM114527.1   R-20821 (T)	---	---	---
P160	<i>P. caeni</i> DSM 24390 (T)	NR_116388.1   HY-14 (T)	---	---	---
P172	<i>P. chlororaphis</i> HT66	KF857481.1   HT66	---	---	---
P187	<i>P. cremoricolorata</i> DSM 17059 (T)	AB060137.1   IAM 1541 (T)	---	---	---
P189	<i>P. cuatrocienegasensis</i> CIP 109853 (T)	NR_044569.1   1N (T)	---	---	---
P198	<i>P. fluorescens</i> 2-79	FJ652606.1   2-79	---	---	---
P232	<i>P. fragi</i> A22	JX184912.1   A22	---	---	---
P233	<i>P. fragi</i> B25 (T)	AB021413.1   IFO 3458 (T)	---	---	---
P237	<i>P. fulva</i> NBRC 16637 (T)	D84015.1   IAM 1529 (T)	---	---	---
P241	<i>P. graminis</i> DSM 11363 (T)	Y11150.1   DSM 11363 (T)	---	---	---
P261	<i>P. marincola</i> JCM 14761 (T)	AB301071.1   KMM 3042 (T)	---	---	---
P273	<i>P. monteilii</i> NBRC 103158 (T)	AF064458.1   CIP 104883 (T)	---	---	---
P277	<i>P. mosselii</i> DSM 17497 (T)	AF072688.2   CIP 105259 (T)	---	---	---
P282	<i>P. oleovorans</i> (ssp. <i>lubricantis</i> ) RS1 (T)	NR_115874.1   RS1 (T)	---	---	---
P283	<i>P. oleovorans</i> (ssp. <i>oleovorans</i> ) NBRC 13583 (T)	D84018.1   IAM 1508 (T)	---	---	---

ID	Stammbezeichner <sub>NCBI</sub>	16S	<i>gyrB</i>	<i>rpoB</i>	<i>rpoD</i>
		Sequenz Stammkennung	Sequenz Stammkennung	Sequenz Stammkennung	Sequenz Stammkennung
P284	<i>P. oleovorans</i> MGY01	KF992130.1   MGY01	---	---	---
P055	<i>P. orientalis</i> DSM 17489 (T)	AF064457.1   CFML 96-170 (T)	---	---	---
P287	<i>P. otitidis</i> DSM 17224 (T)	AY953147.1   MCC10330 (T)	---	---	---
P292	<i>P. parafulva</i> DSM 17004 (T)	AB060132.1   AJ 2129 (T)	---	---	---
P296	<i>P. plecoglossicida</i> NBRC 103162 (T)	AB009457.1   FPC 951 (T)	---	---	---
P308	<i>P. pseudoalcaligenes</i> NBRC 14167	AB021379.1   JCM 5968 (T)	---	---	---
P309	<i>P. psychrophila</i> HA-4	JQ968688.1   HA-4	---	---	---
P328	<i>P. putida</i> PD1	KF443801.1   PD1	---	---	---
P341	<i>P. resinovorans</i> DSM 21078 (T)	AB021373.1   ATCC 14235 (T)	---	---	---
P357	<i>P. soli</i> LMG 27941T	NR_134794.1   F-279,208 (T)	---	---	---
P377	<i>P. stutzeri</i> TS44	EU073110.1   TS44	---	---	---
P424	<i>P. taiwanensis</i> DSM 21245 (T)	NR_116172.1   BCRC 17751 (T)	---	---	---
P425	<i>P. thermotolerans</i> DSM 14292 (T)	NR_028008.1   CM3 (T)	---	---	---
P429	<i>P. toyotomiensis</i> JCM 15604 (T)	NR_112808.1   HT-3 (T)	---	---	---
P437	<i>P. vranovensis</i> DSM 16006 (T)	NR_043313.1   2B2 (T)	---	---	---

**ID**, Kennung zur eindeutigen Identifizierung eines Stammes

**Stammbezeichner<sub>NCBI</sub>**, Bezeichnung des Stammes, wobei Gattungs- und Artzugehörigkeit sowie die Stammkennung vom NCBI übernommen wurden

**Sequenz|Stammkennung**, Kennungen der NCBI oder Genbank Referenzsequenz gefolgt von Stammkennung der Referenzsequenz

(T), anerkannter Typstamm nach LPSN [67] (Stand Apr. 2017)

Tabelle 31 Verwendete *Bacillus cereus* s. l. NCBI Genomdaten

ID	Stammbezeichnung <sub>NCBI</sub>	WGS Kennung	RefSeq/GenBank Sequenzen
B009	<i>B. anthracis</i> 2002013094		NZ_CP009900.1 bis NZ_CP009902.1
B010	<i>B. anthracis</i> 'Ames Ancestor'		NC_007530.2, NC_007322.2, NC_007323.3
B011	<i>B. anthracis</i> BA1035		NZ_CP009698.1 bis NZ_CP009700.1
B012	<i>B. anthracis</i> HYU01		NZ_CP008846.1 bis NZ_CP008848.1
B013	<i>B. anthracis</i> K3		NZ_CP009329.1 bis NZ_CP009331.1
B014	<i>B. anthracis</i> Ohio ACB		NZ_CP009339.1 bis NZ_CP009341.1
B015	<i>B. anthracis</i> Sterne		NZ_CP009540.1, NZ_CP009541.1
B016	<i>B. anthracis</i> Vollum		NZ_CP007664.1 bis NZ_CP007666.1
B017	<i>B. "bingmoyongensis"</i> FJAT-13831 (iT)	AKCS01	
B018	<i>B. cereus</i> #17	JYFW01	
B019	<i>B. cereus</i> 03BB102		NC_012472.1, NC_012473.1
B020	<i>B. cereus</i> 03BB108		NZ_CP009634.1 bis NZ_CP009641.1
B021	<i>B. cereus</i> 03BB87		NZ_CP009939.1 bis NZ_CP009941.1
B022	<i>B. cereus</i> 105Ne	JYCK01	
B023	<i>B. cereus</i> 13061	JMPX01	
B024	<i>B. cereus</i> 14294-3	JYPF01	
B025	<i>B. cereus</i> 172560W	ACLV01	
B026	<i>B. cereus</i> 6/27/S	LABV01	
B027	<i>B. cereus</i> 7/27/S	LABW01	
B028	<i>B. cereus</i> 95/8201	ACMF01	
B029	<i>B. cereus</i> A1	JHOG01	
B030	<i>B. cereus</i> AH1134	ABDA02	
B031	<i>B. cereus</i> AH1271	ACMR01	
B032	<i>B. cereus</i> AH1272	ACMS01	
B033	<i>B. cereus</i> AH1273	ACMT01	
B034	<i>B. cereus</i> AH187		NC_011654.1 bis NC_011658.1
B035	<i>B. cereus</i> AH603	ACMP01	
B036	<i>B. cereus</i> AH621	ACLX01	
B037	<i>B. cereus</i> AH676	ACMQ01	
B038	<i>B. cereus</i> AH820		NC_011771.1, NC_011773.1, NC_011776.1, NC_011777.1
B039	<i>B. cereus</i> AND1407	AHCM01	
B040	<i>B. cereus</i> ATCC 10876	JMPW01	

ID	Stammbezeichnung <sub>NCBI</sub>	WGS Kennung	RefSeq/GenBank Sequenzen
B041	<i>B. cereus</i> ATCC 10987		NC_003909.8, NC_005707.1
B042	<i>B. cereus</i> ATCC 14579 (T)		NC_004721.2, NC_004722.1
B043	<i>B. cereus</i> ATCC 4342		NZ_CP009627.1, NZ_CP009628.1
B044	<i>B. cereus</i> B4077	LCYI01	
B045	<i>B. cereus</i> B4078	LCYJ01	
B046	<i>B. cereus</i> B4080	LCYK01	
B047	<i>B. cereus</i> B4086	LCYL01	
B048	<i>B. cereus</i> B4087	LCYM01	
B049	<i>B. cereus</i> B4147	LCYN01	
B050	<i>B. cereus</i> B4153	LCYO01	
B051	<i>B. cereus</i> B4158	LCYP01	
B052	<i>B. cereus</i> B4264		NC_011725.1
B053	<i>B. cereus</i> B5-2	AHFW01	
B054	<i>B. cereus</i> BAG10-1	AHCN01	
B055	<i>B. cereus</i> BAG10-2	AHCO01	
B056	<i>B. cereus</i> BAG10-3	AHCP01	
B057	<i>B. cereus</i> BAG1X1-1	AHCQ01	
B058	<i>B. cereus</i> BAG1X1-2	AHCR01	
B059	<i>B. cereus</i> BAG1X1-3	AHCS01	
B060	<i>B. cereus</i> BAG1X2-2	AHCU01	
B061	<i>B. cereus</i> BAG1X2-3	AHCV01	
B062	<i>B. cereus</i> BAG20-1	AHCW01	
B063	<i>B. cereus</i> BAG20-2	AHCX01	
B064	<i>B. cereus</i> BAG20-3	AHCY01	
B065	<i>B. cereus</i> BAG2X1-1	AHCZ01	
B066	<i>B. cereus</i> BAG2X1-2	AHDA01	
B067	<i>B. cereus</i> BAG2X1-3	AHDB01	
B068	<i>B. cereus</i> BAG30-1	AHJV01	
B069	<i>B. cereus</i> BAG30-2	AHDC01	
B070	<i>B. cereus</i> BAG3X2-1	AHDD01	
B071	<i>B. cereus</i> BAG3X2-2	AHDE01	
B072	<i>B. cereus</i> BAG40-1	AHDF01	
B073	<i>B. cereus</i> BAG4X12-1	AHDG01	

ID	Stammbezeichnung <sub>NCBI</sub>	WGS Kennung	RefSeq/GenBank Sequenzen
B074	<i>B. cereus</i> BAG4X2-1	AHDH01	
B075	<i>B. cereus</i> BAG5O-1	AHDI01	
B076	<i>B. cereus</i> BAG5X1-1	AHDJ01	
B077	<i>B. cereus</i> BAG5X2-1	AHDL01	
B078	<i>B. cereus</i> BAG6O-1	AHDM01	
B079	<i>B. cereus</i> BAG6O-2	AHDN01	
B080	<i>B. cereus</i> BAG6X1-1	AHDO01	
B081	<i>B. cereus</i> BAG6X1-2	AHDP01	
B082	<i>B. cereus</i> BcFL2013	JHQN01	
B083	<i>B. cereus</i> BDRD-Cer4	ACME01	
B084	<i>B. cereus</i> BDRD-ST24	ACMB01	
B085	<i>B. cereus</i> BDRD-ST26	ACMC01	
B086	<i>B. cereus</i> BGSC 6E1	ACLU01	
B087	<i>B. cereus</i> BMG1.7	AHDQ01	
B088	<i>B. cereus</i> BtB2-4	AHDR01	
B089	<i>B. cereus</i> CI		NC_014331.1 bis NC_014333.1, NC_014335.1
B090	<i>B. cereus</i> D17		NZ_CP009299.1, NZ_CP009300.1
B091	<i>B. cereus</i> E33L		NC_006274.1, NC_007103.1 bis NC_007107.1
B092	<i>B. cereus</i> F	AHHI01	
B093	<i>B. cereus</i> F1-15	JMSG01	
B094	<i>B. cereus</i> F3162-04	JZEV01	
B095	<i>B. cereus</i> F3175-03	JYPI01	
B096	<i>B. cereus</i> F4429-71	JYPJ01	
B097	<i>B. cereus</i> F528-94	JYPH01	
B098	<i>B. cereus</i> F65185	ACMO01	
B099	<i>B. cereus</i> F837/76		NC_016779.1, NC_016780.1, NC_016794.1
B100	<i>B. cereus</i> FM1		NZ_CP009368.1, NZ_CP009369.1
B101	<i>B. cereus</i> FORC_005		NZ_CP009686.1
B102	<i>B. cereus</i> FRI-35		NC_018491.1 bis NC_018494.1, NC_018499.1
B103	<i>B. cereus</i> G9241		NZ_CP009589.1 bis NZ_CP009592.1
B104	<i>B. cereus</i> G9842		NC_011772.1, NC_011774.1, NC_011775.1
B105	<i>B. cereus</i> HD73	AH DU01	
B106	<i>B. cereus</i> HuA2-1	AH DV01	

ID	Stammbezeichnung <sub>NCBI</sub>	WGS Kennung	RefSeq/GenBank Sequenzen
B107	<i>B. cereus</i> HuA2-3	AHDW01	
B108	<i>B. cereus</i> HuA2-4	AHDX01	
B109	<i>B. cereus</i> HuA2-9	AHDY01	
B110	<i>B. cereus</i> HuA3-9	AHDZ01	
B111	<i>B. cereus</i> HuA4-10	AHEA01	
B112	<i>B. cereus</i> HuB1-1	AHEB01	
B113	<i>B. cereus</i> HuB13-1	AHEC01	
B114	<i>B. cereus</i> HuB2-9	AHED01	
B115	<i>B. cereus</i> HuB4-10	AHEE01	
B116	<i>B. cereus</i> HuB4-4	AHEF01	
B117	<i>B. cereus</i> HuB5-5	AHEG01	
B118	<i>B. cereus</i> HWW 274-2	JYPL01	
B119	<i>B. cereus</i> INRA A3	LABH01	
B120	<i>B. cereus</i> INRA C3	LABI01	
B121	<i>B. cereus</i> IS195	AHEH01	
B122	<i>B. cereus</i> ISP2954	AHEJ01	
B123	<i>B. cereus</i> ISP3191	AHEK01	
B124	<i>B. cereus</i> KS100Bc	LBBY01	
B125	<i>B. cereus</i> LCT-BC244	AJGQ01	
B126	<i>B. cereus</i> LK9	LDUP01	
B127	<i>B. cereus</i> Lr 4/2	JYCM01	
B128	<i>B. cereus</i> m1293	ACLS01	
B129	<i>B. cereus</i> m1550	ACMA01	
B130	<i>B. cereus</i> MC118	AHEM01	
B131	<i>B. cereus</i> MC67	AHEN01	
B132	<i>B. cereus</i> MHI 226	LABL01	
B133	<i>B. cereus</i> MHI 86	LABK01	
B134	<i>B. cereus</i> MIT0214	JXDH01	
B135	<i>B. cereus</i> MM3	ACLW01	
B136	<i>B. cereus</i> MSX-A1	AHEO01	
B137	<i>B. cereus</i> MSX-A12	AHEP01	
B138	<i>B. cereus</i> MSX-D12	AHEQ01	
B139	<i>B. cereus</i> NC7401		NC_016771.1 bis NC_016774.1, NC_016792.1, NC_016793.1

ID	Stammbezeichnung <sub>NCBI</sub>	WGS Kennung	RefSeq/GenBank Sequenzen
B140	<i>B. cereus</i> NJ-W		NZ_CP012483.1 bis NZ_CP012486.1
B141	<i>B. cereus</i> NVH 0075-95	LABM01	
B142	<i>B. cereus</i> NVH0597-99	ABDK02	
B143	<i>B. cereus</i> Q1		NC_011969.1, NC_011971.1, NC_011973.1
B144	<i>B. cereus</i> R309803	ACLY01	
B145	<i>B. cereus</i> RIMV BC 126	LABO01	
B146	<i>B. cereus</i> RIVM BC 90	LABN01	
B147	<i>B. cereus</i> RIVM BC 934	LABP01	
B148	<i>B. cereus</i> RIVM BC 964	LABQ01	
B149	<i>B. cereus</i> Rock1-15	ACMH01	
B150	<i>B. cereus</i> Rock1-3	ACMG01	
B151	<i>B. cereus</i> Rock3-28		NZ_CM000730.1
B152	<i>B. cereus</i> Rock3-29	ACMJ01	
B153	<i>B. cereus</i> Rock3-42	ACMK01	
B154	<i>B. cereus</i> Rock3-44		NZ_CM000733.1
B155	<i>B. cereus</i> Rock4-18	ACMN01	
B156	<i>B. cereus</i> Rock4-2	ACMM01	
B157	<i>B. cereus</i> S2-8		NZ_CP009603.1 bis NZ_CP009606.1
B158	<i>B. cereus</i> S612	LJXA01	
B159	<i>B. cereus</i> SDA KA 96	LABR01	
B160	<i>B. cereus</i> TIAC219	AHCJ01	
B161	<i>B. cereus</i> tsu1	JPYN01	
B162	<i>B. cereus</i> VD014	AHER01	
B163	<i>B. cereus</i> VD021	AHES01	
B164	<i>B. cereus</i> VD045	AHET01	
B165	<i>B. cereus</i> VD048	AHEU01	
B166	<i>B. cereus</i> VD078	AHEV01	
B167	<i>B. cereus</i> VD102	AHEW01	
B168	<i>B. cereus</i> VD107	AHEX01	
B169	<i>B. cereus</i> VD115	AHEY01	
B170	<i>B. cereus</i> VD118	AHEZ01	
B171	<i>B. cereus</i> VD131	AHFA01	
B172	<i>B. cereus</i> VD133	AHFB01	

ID	Stammbezeichnung <sub>NCBI</sub>	WGS Kennung	RefSeq/GenBank Sequenzen
B173	<i>B. cereus</i> VD136	AHFC01	
B174	<i>B. cereus</i> VD140	AHFD01	
B175	<i>B. cereus</i> VD142	AHCL02	
B176	<i>B. cereus</i> VD146	AHFE01	
B177	<i>B. cereus</i> VD148	AHFF01	
B178	<i>B. cereus</i> VD154	AHFG01	
B179	<i>B. cereus</i> VD156	AHFH01	
B180	<i>B. cereus</i> VD166	AHFI01	
B181	<i>B. cereus</i> VD169	AHFJ01	
B182	<i>B. cereus</i> VD184	AHFK01	
B183	<i>B. cereus</i> VD196	AHFL01	
B184	<i>B. cereus</i> VD200	AHFM01	
B185	<i>B. cereus</i> VD214	AHFN01	
B186	<i>B. cereus</i> VDM006	AHFT01	
B187	<i>B. cereus</i> VDM019	AHFO01	
B188	<i>B. cereus</i> VDM021	AHFU01	
B189	<i>B. cereus</i> VDM022	AHFP01	
B190	<i>B. cereus</i> VDM034	AHFQ01	
B191	<i>B. cereus</i> VDM053	AHFR01	
B192	<i>B. cereus</i> VDM062	AHFS01	
B193	<i>B. cereus</i> W	ABCZ02	
B194	<i>B. cereus</i> WSBC 10035	LABS01	
B195	<i>B. cytotoxicus</i> CVUAS 2833	JYPG01	
B196	<i>B. cytotoxicus</i> NVH 391-98 (T)		NC_009673.1, NC_009674.1
B197	<i>B. "gaemokensis"</i> JCM 15801 (iT)	JOTM01	
B198	<i>B. "manliponensis"</i> JCM 15802 (iT)	JOTN01	
B199	<i>B. mycoides</i> ATCC 6462 (T)		NZ_CP009689.1 bis NZ_CP009692.1
B200	<i>B. mycoides</i> B38V	JYFS01	
B201	<i>B. mycoides</i> BHP	JMQC01	
B202	<i>B. mycoides</i> Flugge 10206	JMQD01	
B203	<i>B. mycoides</i> Rock3-17	ACMW01	
B204	<i>B. mycoides</i> WSBC 10969	LABT01	
B205	<i>B. pseudomycoides</i> DSM 12442 (T)		NZ_CM000745.1

ID	Stammbezeichnung <sub>NCBI</sub>	WGS Kennung	RefSeq/GenBank Sequenzen
B206	<i>B. sp.</i> 105MF	ARGR01	
B207	<i>B. sp.</i> 123MFChir2	ARSX01	
B208	<i>B. sp.</i> 7_6_55CFAA_CT2	ACWE01	
B209	<i>B. sp.</i> 95MFCvi2.1	ARGV01	
B210	<i>B. sp.</i> GeD10	CAVI01	
B211	<i>B. sp.</i> H1a	AYMH01	
B212	<i>B. sp.</i> H1m	AYMM01	
B213	<i>B. sp.</i> L_1B0_5	JXIS01	
B214	<i>B. sp.</i> LK2	LDUK01	
B215	<i>B. sp.</i> Root11	LMJP01	
B216	<i>B. sp.</i> Root131	LMJQ01	
B217	<i>B. sp.</i> UMTAT18	JSFD01	
B218	<i>B. sp.</i> UNC322MFChir4.1	JQKQ01	
B219	<i>B. sp.</i> UNC437CL72CviS29	AXVB01	
B220	<i>B. thuringiensis</i> 147	LFXM01	
B221	<i>B. thuringiensis</i> 27	JWJY01	
B222	<i>B. thuringiensis</i> 4Q7	JEOC01	
B223	<i>B. thuringiensis</i> 823_BTHU	JURK01	
B224	<i>B. thuringiensis</i> 97-27		NZ_CP010087.1, NZ_CP010088.1
B225	<i>B. thuringiensis</i> AK47	JWST01	
B226	<i>B. thuringiensis</i> Al Hakam		NZ_CP009645.1 bis NZ_CP009651.1
B227	<i>B. thuringiensis</i> ATCC 10792 (T)	ACNF01	
B228	<i>B. thuringiensis</i> BAC3151	LDKD01	
B229	<i>B. thuringiensis</i> BGSC 4AA1		NZ_CP010577.1 bis NZ_CP010583.1
B230	<i>B. thuringiensis</i> BGSC 4AJ1	ACNE01	
B231	<i>B. thuringiensis</i> BGSC 4AW1	ACNG01	
B232	<i>B. thuringiensis</i> BGSC 4BD1	ACNI01	
B233	<i>B. thuringiensis</i> BGSC 4CC1	ACNJ01	
B234	<i>B. thuringiensis</i> BGSC 4Y1	ACMY01	
B235	<i>B. thuringiensis</i> BMB171		NC_014171.1, NC_014172.1
B236	<i>B. thuringiensis</i> BMB3201	LDEU01	
B237	<i>B. thuringiensis</i> BR58	LIIT01	
B238	<i>B. thuringiensis</i> Bt407		NC_018877.1 bis NC_018886.1

ID	Stammbezeichnung <sub>NCBI</sub>	WGS Kennung	RefSeq/GenBank Sequenzen
B239	<i>B. thuringiensis</i> CT-43		NC_017202.1 bis NC_017212.1
B240	<i>B. thuringiensis</i> DAR 81934	ANPK01	
B241	<i>B. thuringiensis</i> DB27	CBXL01	
B242	<i>B. thuringiensis</i> HD 600	JTHH01	
B243	<i>B. thuringiensis</i> HD-1		NZ_CP004870.1 bis NZ_CP004883.1
B244	<i>B. thuringiensis</i> HD1002		NZ_CP009344.1 bis NZ_CP009351.1
B245	<i>B. thuringiensis</i> HD1011		NZ_CP009332.1 bis NZ_CP009336.1
B246	<i>B. thuringiensis</i> HD-29		NZ_CP010089.1 bis NZ_CP010099.1
B247	<i>B. thuringiensis</i> HD521		NZ_CP010106.1 bis NZ_CP010112.1
B248	<i>B. thuringiensis</i> HD571		NZ_CP009599.1, NZ_CP009600.1
B249	<i>B. thuringiensis</i> HD682		NZ_CP009717.1 bis NZ_CP009720.1
B250	<i>B. thuringiensis</i> HD73		NC_020238.1 bis NC_020243.1, NC_020249.1, NC_020250.1
B251	<i>B. thuringiensis</i> HD-771		NC_018486.1 bis NC_018490.1, NC_018500.1 bis NC_018503.1
B252	<i>B. thuringiensis</i> HD-789		NC_018508.1 bis NC_018512.1, NC_018516.1, NC_018517.1
B253	<i>B. thuringiensis</i> HS18-1		NZ_CP012099.1 bis NZ_CP012108.1
B254	<i>B. thuringiensis</i> Hu4-2	AMXT02	
B255	<i>B. thuringiensis</i> IBL 200	ACNK01	
B256	<i>B. thuringiensis</i> IBL 4222	ACNL01	
B257	<i>B. thuringiensis</i> IS5056		NC_020376.1 bis NC_020385.1, NC_020390.1 bis NC_020394.1
B258	<i>B. thuringiensis</i> JM-Mgvxx-63	AYSM01	
B259	<i>B. thuringiensis</i> Leapi01	AMXS02	
B260	<i>B. thuringiensis</i> LM1212	AYPV01	
B261	<i>B. thuringiensis</i> Lr3/2	JYCH01	
B262	<i>B. thuringiensis</i> Lr7/2	JYCL01	
B263	<i>B. thuringiensis</i> MC28		NC_018684.1 bis NC_018689.1, NC_018693.1, NC_018694.1
B264	<i>B. thuringiensis</i> NA205-2	AYXQ01	
B265	<i>B. thuringiensis</i> NBIN-866	JEOF01	
B266	<i>B. thuringiensis</i> Sbt003	AMYJ01	
B267	<i>B. thuringiensis</i> T01001	ACNA01	
B268	<i>B. thuringiensis</i> T01-328	ARXZ02	
B269	<i>B. thuringiensis</i> T03a001	ACND01	
B270	<i>B. thuringiensis</i> T13001		NZ_CM000750.1
B271	<i>B. thuringiensis</i> YBT-020		NC_017199.1 bis NC_017201.1

ID	Stammbezeichner <sub>NCBI</sub>	WGS Kennung	RefSeq/GenBank Sequenzen
B272	<i>B. thuringiensis</i> YBT-1518		NC_020124.1, NC_022873.1 bis NC_022877.1, NC_022882.1
B273	<i>B. thuringiensis</i> YBT-1520		NZ_CP004858.1 bis NZ_CP004869.1
B274	<i>B. thuringiensis</i> YC-10		NZ_CP011349.1 bis NZ_CP011358.1
B275	<i>B. thuringiensis</i> ZBG4	LFQM01	
B276	<i>B. toyonensis</i> BCT-7112 (T)		NC_022781.1 bis NC_022783.1
B277	<i>B. weihenstephanensis</i> FSL H7-687	ASPY01	
B278	<i>B. weihenstephanensis</i> FSL R5-860	ASPZ01	
B279	<i>B. weihenstephanensis</i> JAS 83/3	JNLY01	
B280	<i>B. weihenstephanensis</i> KBAB4		NC_010180.1 bis NC_010184.1
B281	<i>B. weihenstephanensis</i> WSBC 10204 (T)		NZ_CP009746.1

**ID**, Kennung zur eindeutigen Identifizierung eines Stammes

**Stammbezeichner<sub>NCBI</sub>**, Bezeichnung des Stammes, wobei Gattungs- und Artzugehörigkeit sowie die Stammkennung vom NCBI übernommen wurden

**WGS Kennung**, Kennung des WGS Projektes bestehend aus einem 4 buchstabigen Präfix (Projekt) gefolgt von 2 Ziffern (Versionsnummer)

**RefSeq/Genbank Sequenzen**, Kennungen der NCBI oder Genbank Referenzgenomsequenzen

(T), anerkannter Typstamm nach LPSN [67]

(iT), nicht anerkannter Typstamm, da Arname (noch) nicht in LPSN [67] aufgenommen

Tabelle 32 Verworfenne *Bacillus* NCBI Genomdaten

Stammbezeichnung <sup>NCBI</sup>	WGS Kennung	RefSeq/GenBank Sequenzen	Grund des Ausschlusses
<i>B. cereus</i> 3a		NZ_CP009593.1 bis NZ_CP009596.1	AF ≥ 0,99 und gANI ≥ 99,99 zu B156
<i>B. cereus</i> 822_BCER	JURL01		AF ≥ 0,99 und gANI ≥ 99,99 zu B223
<i>B. cereus</i> BAG1X2-1	AHCT01		AF ≥ 0,99 und gANI ≥ 99,99 zu B056
<i>B. cereus</i> BAG5X12-1	AHDK01		AF ≥ 0,99 und gANI ≥ 99,99 zu B072
<i>B. cereus</i> BDRD-ST196	ACMD01		AF ≥ 0,99 und gANI ≥ 99,99 zu B281
<i>B. cereus</i> CER057	AHDS01		AF ≥ 0,99 und gANI ≥ 99,99 zu B087
<i>B. cereus</i> CER074	AHDT01		AF ≥ 0,99 und gANI ≥ 99,99 zu B087
<i>B. cereus</i> IP5832	LABJ01		AF ≥ 0,99 und gANI ≥ 99,99 zu B276
<i>B. cereus</i> IS075	AHCH02		AF ≥ 0,99 und gANI ≥ 99,99 zu B120
<i>B. cereus</i> IS845/00	AHEI01		AF ≥ 0,99 und gANI ≥ 99,99 zu B120
<i>B. cereus</i> LCT-BC235	ATHN01		AF ≥ 0,99 und gANI ≥ 99,99 zu B124
<i>B. cereus</i> LCT-BC25	ATHM01		AF ≥ 0,99 und gANI ≥ 99,99 zu B124
<i>B. cereus</i> Schrouff	AHCI01		AF ≥ 0,99 und gANI ≥ 99,99 zu <i>B. cereus</i> K-5975c
<i>B. cereus</i> VD022	AHCK01		AF ≥ 0,99 und gANI ≥ 99,99 zu B159
<i>B. mycooides</i> 219298		NZ_CP007621.1 bis NZ_CP007626.1	AF ≥ 0,99 und gANI ≥ 99,99 zu B226
<i>B. thuringiensis</i> AlHakam		NC_008598.1, NC_008600.1	AF ≥ 0,99 und gANI ≥ 99,99 zu B248
<i>B. thuringiensis</i> BGSC 4BA1	ACNH01		AF ≥ 0,99 und gANI ≥ 99,99 zu B245
<i>B. cereus</i> 03BB102		NZ_CP009317.1, NZ_CP009318.1	Duplikat zu B018
<i>B. cereus</i> 03BB108	ABDM02		Duplikat zu B019
<i>B. cereus</i> ATCC 10876	ACLT01		Duplikat zu B039
<i>B. cereus</i> ATCC 4342	JMPY01		Duplikat zu B042
<i>B. cereus</i> ATCC 4342	ACLZ01		Duplikat zu B042
<i>B. cereus</i> E33L		NZ_CP009965.1 bis NZ_CP009970.1	Duplikat zu B090
<i>B. cereus</i> G9241	AAEK01		Duplikat zu B102
<i>B. mycooides</i> DSM 2048 (T)	ACMU01		Duplikat zu B198
<i>B. thuringiensis</i> 97-27		NC_005957.1, NC_006578.1	Duplikat zu B224
<i>B. thuringiensis</i> Bt407	ACMZ01		Duplikat zu B238
<i>B. thuringiensis</i> HD-1		NZ_CP009998.1, NZ_CP009999.1, NZ_CP010000.1 bis NZ_CP010012.1	Duplikat zu B243
<i>B. thuringiensis</i> HD-1	JMHW01		Duplikat zu B243
<i>B. thuringiensis</i> YBT-1520		NZ_CP007607.1 bis NZ_CP007616.1	Duplikat zu B273
<i>B. weihenstephanensis</i> NBRC 101238 (T)	BAUY01		Duplikat zu B281
<i>B. mycooides</i> 11kri323	JYCJ01		Zu hoher GC-Gehalt von 38,2 %

Stammbezeichner <sub>NCBI</sub>	WGS Kennung	RefSeq/GenBank Sequenzen	Grund des Ausschlusses
<i>B. thuringiensis</i> 1094BTHU	JWDE01		Zu niedriger GC-Gehalt von 29,7 % und zu geringe Assembly Size $\approx$ 1.9 Mpb
<i>B. sp.</i> 522_BSPC	JVDT01		Maximale AF < 0,25
<i>B. sp.</i> FF3	CCAP01		Maximale AF < 0,25
<i>B. sp.</i> FJAT-22090		NZ_CP012601.1	Maximale AF < 0,25
<i>B. sp.</i> FJAT-25496	LMBY01		Maximale AF < 0,25
<i>B. sp.</i> J37	JAEK01		Maximale AF < 0,25
<i>B. sp.</i> KW-12	CCFJ01		Maximale AF < 0,25
<i>B. sp.</i> KW-12	CCNM01		Maximale AF < 0,25
<i>B. sp.</i> MB2021	JNJJ01		Maximale AF < 0,25
<i>B. sp.</i> REN51N	JXAB01		Maximale AF < 0,25
<i>B. sp.</i> SA1-12	LATZ01		Maximale AF < 0,25
<i>B. sp.</i> SIT1	CWJM01		Maximale AF < 0,25
<i>B. sp.</i> TS-2	BAWL01		Maximale AF < 0,25
<i>B. cereus</i> H3081.97	ABDL02		Kein <i>adk</i> Gen
<i>B. cereus</i> K-5975c	AHEL01		Kein <i>pycA</i> Gen
<i>B. mycooides</i> Rock1-4		CM000743.1	Unvollständiges <i>panC</i> Gen

**Stammbezeichner<sub>NCBI</sub>**, Bezeichnung des Stammes, wobei Gattungs- und Artzugehörigkeit sowie die Stammkennung vom NCBI übernommen wurden

**WGS Kennung**, Kennung des WGS Projektes bestehend aus einem 4 buchstabigen Präfix (Projekt) gefolgt von 2 Ziffern (Versionsnummer)

**RefSeq/Genbank Sequenzen**, Kennungen der NCBI oder Genbank Referenzgenomsequenzen

**Grund des Ausschlusses**, Grund weshalb die NCBI Genom Assemblierungen keine weitere Verwendung fanden

(T), anerkannter Typstamm nach LPSN [67]

Tabelle 33 NCBI Protein Referenzsequenzen untersuchter *B. cereus s.l.* Virulenzfaktoren (ohne Bt-Toxine)

Virulenzfaktor	RefSeq/GenBank Sequenz
<b>Klasse: Enterotoxin-Komplexe</b>	
NheA	NheA <sub>a</sub> ABY42976.1 ABY46573.1
NheB	NheB <sub>a</sub> ABY42977.1 ABY46572.1
NheC	NheC <sub>a</sub> ABY42978.1 ABY46571.1
HblC	HblC <sub>a</sub> WP_000850237.1 WP_001996176.1
HblD	HblD <sub>a</sub> WP_000714433.1 WP_001996093.1
HblA	HblA <sub>a</sub> WP_000591961.1 WP_001996202.1
HblB	WP_001994017.1
<b>Klasse: Hämolysine</b>	
CytK	NP_830896.1
HlyII	NP_833256.1
HlyIII	NP_831961.1
HlyA	NP_833889.1
<b>Klasse: Metalloproteasen</b>	
InhA1	NP_831066.1
InhA2	NP_830479.1
InhA3	NP_832730.1
<b>Klasse: Anthrax-Toxin</b>	
EF	WP_000197748.1
PA	WP_000746486.1
LF	WP_001022097.1
<b>Bikarbonat-Transporter</b>	
ATP-Bindeprotein	WP_000218630.1
Permease	WP_000432120.1
Substrat-Bindeprotein	WP_000276941.1
<b>Klasse: Proteine des Cereulid-Synthetase-Genclusters</b>	
CesH	WP_032491695.1
CesP	WP_000680399.1
CesT	WP_000764755.1
CesA	WP_002081542.1
CesB	WP_000953496.1
CesC	WP_000590108.1
CesD	WP_001008264.1

Tabelle 34 NCBI Protein Referenzsequenzen untersuchter Bt-Toxine

Holotyp	RefSeq/GenBank Sequenz						
<b>Cry</b>		<b>Cry</b>		<b>Cry</b>		<b>Cry</b>	
Cry1Aa1	AAA22353.1	Cry5Ad1	ABQ82087.1	Cry13Aa1	AAA22356.1	Cry32Wa1	WP_029439213.1
Cry1Ab1	AAA22330.1	Cry5Ba1	AAA68598.1	Cry14Aa1	AAA21516.1	Cry33Aa1	AAL26871.1
Cry1Ac1	AAA22331.1	Cry5Ca1	AEB52309.1	Cry14Ab1	AGU13817.1	Cry34Aa1	AAG50341.1
Cry1Ad1	AAA22340.1	Cry5Da1	AEB52310.1	Cry15Aa1	AAA22333.1	Cry34Ab1	AAG41671.1
Cry1Ae1	AAA22410.1	Cry5Ea2	ZP_04124038.1	Cry16Aa1	CAA63860.1	Cry34Ac1	AAG50118.1
Cry1Af1	AAB82749.1	Cry6Aa1	AAA22357.1	Cry17Aa1	CAA67841.1	Cry34Ba1	AAK64565.1
Cry1Ag1	AAD46137.1	Cry6Ba1	AAA22358.1	Cry18Aa1	CAA67506.1	Cry35Aa1	AAG50342.1
Cry1Ah1	AAQ14326.1	Cry7Aa1	AAA22351.1	Cry18Ba1	AAF89667.1	Cry35Ab1	AAG41672.1
Cry1Ai1	AAO39719.1	Cry7Ab1	AAA21120.1	Cry18Ca1	AAF89668.1	Cry35Ac1	AAG50117.1
Cry1Ba1	CAA29898.1	Cry7Ba1	ABB70817.1	Cry19Aa1	CAA68875.1	Cry35Ba1	AAK64566.1
Cry1Bb1	AAA22344.1	Cry7Bb1	AGU13818.1	Cry19Ba1	BAA32397.1	Cry36Aa1	AAK64558.1
Cry1Bc1	CAA86568.1	Cry7Ca1	ABR67863.1	Cry19Ca1	AFM37572.1	Cry37Aa1	AAF76376.1
Cry1Bd1	AAD10292.1	Cry7Cb1	AGU13863.1	Cry20Aa1	AAB93476.1	Cry38Aa1	AAK64559.1
Cry1Be1	AAC32850.1	Cry7Da1	ACQ99547.1	Cry20Ba1	ACS93601.1	Cry39Aa1	BAB72016.2
Cry1Bf1	CAC50778.1	Cry7Ea1	AEA92301.1	Cry21Aa1	I32932.1	Cry40Aa1	BAB72018.1
Cry1Bg1	AAO39720.1	Cry7Fa1	AEA92303.1	Cry21Ba1	BAC06484.1	Cry40Ba1	BAC77648.1
Cry1Bh1	ADT78693.1	Cry7Fb2	AGU13847.1	Cry21Ca1	AEH76817.1	Cry40Ca1	ACB14343.1
Cry1Bi1	AGU13865.1	Cry7Ga2	AGU13834.1	Cry21Da1	AEH76818.1	Cry40Da1	ACF15199.1
Cry1Bj1	AMB62179.1	Cry7Gb1	AGU13815.1	Cry21Fa1	AHI15915.1	Cry41Aa1	BAD35157.1
Cry1Ca1	CAA30396.1	Cry7Gc1	AGU13819.1	Cry21Ga1	AHI15916.1	Cry41Ab1	BAD35163.1
Cry1Cb2	AAG35409.1	Cry7Gd1	AGU13862.1	Cry21Ha1	AHI15917.1	Cry41Ba1	AEB52311.1
Cry1Da1	CAA38099.1	Cry7Ha1	AGU13816.1	Cry22Aa2	CAD43579.1	Cry41Ca1	WP_033694850.1
Cry1Db1	CAA80234.1	Cry7Ia1	AGU13830.1	Cry22Ab1	AAK50456.1	Cry42Aa1	BAD35166.1
Cry1Dc1	ABK35074.1	Cry7Ja1	AGU13836.1	Cry22Ba1	CAD43578.1	Cry43Aa1	BAD15301.1
Cry1Ea1	CAA37933.1	Cry7Ka1	AGU13845.1	Cry22Bb1	AGU13837.1	Cry43Ba1	BAD15303.1
Cry1Eb1	AAA22346.1	Cry7Kb1	BAM99306.1	Cry23Aa1	AAF76375.1	Cry43Ca1	AGU13841.1
Cry1Fa1	AAA22348.1	Cry7La1	BAM99307.1	Cry24Aa1	AAC61891.1	Cry43Cb1	AGU13860.1
Cry1Fb1	CAA80235.1	Cry8Aa1	AAA21117.1	Cry24Ba1	BAD32657.1	Cry43Cc1	AGU13861.1
Cry1Ga1	CAA80233.1	Cry8Ab1	ABW08088.1	Cry24Ca1	CAJ43600.1	Cry44Aa1	BAD08532.1
Cry1Gb1	AAD10291.1	Cry8Ac1	AGU13827.1	Cry25Aa1	AAC61892.1	Cry45Aa1	BAD22577.1
Cry1Gc1	AAQ52381.1	Cry8Ad1	AGU13849.1	Cry26Aa1	AAD25075.1	Cry45Ba1	WP_029440439.1

Holotyp	RefSeq/GenBank Sequenz	Holotyp	RefSeq/GenBank Sequenz	Holotyp	RefSeq/GenBank Sequenz	Holotyp	RefSeq/GenBank Sequenz
Cry1Ha1	CAA80236.1	Cry8Ba1	AAA21118.1	Cry27Aa1	BAA82796.1	Cry46Aa1	BAC79010.1
Cry1Hb1	AAA79694.1	Cry8Bb1	CAD57542.1	Cry28Aa1	AAD24189.1	Cry46Ab1	BAD35170.1
Cry1Ia1	CAA44633.1	Cry8Bc1	CAD57543.1	Cry29Aa1	CAC80985.1	Cry47Aa1	AAy24695.1
Cry1Ib1	AAA82114.1	Cry8Ca1	AAA21119.1	Cry30Aa1	CAC80986.1	Cry48Aa1	CAJ18351.1
Cry1Ic1	AAC62933.1	Cry8Da1	BAC07226.1	Cry30Ba1	BAD00052.1	Cry48Ab1	CAJ86548.1
Cry1Id1	AAD44366.1	Cry8Db1	BAF93483.1	Cry30Ca1	BAD67157.1	Cry49Aa1	CAH56541.1
Cry1Ie1	AAG43526.1	Cry8Ea1	AAQ73470.1	Cry30Da1	ABN49951.1	Cry49Ab1	CAJ86542.1
Cry1If1	AAQ52382.1	Cry8Fa1	AAT48690.1	Cry30Db1	BAE80088.1	Cry50Aa1	BAE86999.1
Cry1Ig1	AGU13866.1	Cry8Ga1	AAT46073.1	Cry30Ea1	ACC95445.1	Cry50Ba1	ADV57668.1
Cry1Ja1	AAA22341.1	Cry8Ha1	AAW81032.2	Cry30Fa1	ACI22625.1	Cry51Aa1	ABI14444.1
Cry1Jb1	AAA98959.1	Cry8Hb1	AJK90391.1	Cry30Ga1	ACG60020.1	Cry52Aa1	ABU96490.1
Cry1Jc1	AAC31092.1	Cry8Ia1	ACB14342.1	Cry31Aa1	BAB11757.1	Cry52Ba1	ACP43735.1
Cry1Jd1	CAC50779.1	Cry8Ib1	ADT65199.1	Cry31Ab1	BAE79809.1	Cry53Aa1	ABV55105.1
Cry1Ka1	AAB00376.1	Cry8Ja1	ACF37208.1	Cry31Ac1	BAF34368.1	Cry53Ab1	ACP43734.1
Cry1La1	AAS60191.1	Cry8Ka1	ACQ99188.1	Cry31Ad1	BAI44022.1	Cry54Aa1	ACA52194.1
Cry1Ma1	ACT34172.2	Cry8Kb1	ADL27441.1	Cry32Aa1	AAG36711.1	Cry54Ab1	AGC24275.1
Cry1Na1	AGU13813.1	Cry8La1	ADT65198.1	Cry32Ab1	ADO32760.1	Cry54Ba1	ADV57670.1
Cry1Nb1	AGU13843.1	Cry8Ma1	AEB52298.1	Cry32Ba1	BAB78601.1	Cry55Aa1	ABW88932.1
Cry2Aa1	AAA22335.1	Cry8Na1	ADK95169.1	Cry32Ca1	BAB78602.1	Cry56Aa1	ACU57499.1
Cry2Ab1	AAA22342.1	Cry8Pa1	ADV72540.1	Cry32Cb1	AGU13873.1	Cry57Aa1	ACN87261.1
Cry2Ac1	CAA40536.1	Cry8Qa1	ADW78573.1	Cry32Da1	BAB78603.1	Cry58Aa1	ACN87260.1
Cry2Ad1	AAF09583.1	Cry8Ra1	AFP87548.1	Cry32Ea1	ADK66923.1	Cry59Ba1	AEX56526.1
Cry2Ae1	AAQ52362.1	Cry8Ta1	AGU13838.1	Cry32Eb1	AGU13828.1	Cry59Aa1	ACR43758.2
Cry2Af1	ABO30519.1	Cry9Aa1	CAA41122.1	Cry32Fa1	AGU13821.1	Cry60Aa1	ACU24782.1
Cry2Ag1	ACH91610.1	Cry9Ba1	CAA52927.1	Cry32Ga1	AGU13822.1	Cry60Ba1	ADE27985.1
Cry2Ah1	ACL13555.1	Cry9Bb1	AAV28716.1	Cry32Ha1	AGU13826.1	Cry61Aa1	AEA92302.1
Cry2Ai1	ACV97158.1	Cry9Ca1	CAA85764.1	Cry32Hb1	AGU13831.1	Cry62Aa1	ADK91079.1
Cry2Ak1	AGU13825.1	Cry9Da1	BAA19948.1	Cry32Ia1	AGU13832.1	Cry63Aa1	BAI44028.1
Cry2Al1	AHH35153.1	Cry9Db1	AAX78439.1	Cry32Ja1	AGU13850.1	Cry64Aa1	BAJ05397.1
Cry2Ba1	AGU13823.1	Cry9Dc1	AGU13848.1	Cry32Ka1	AGU13853.1	Cry64Ba1	AGT29559.1
Cry3Aa1	AAA22336.1	Cry9Ea1	BAA34908.1	Cry32La1	AGU13854.1	Cry64Ca1	AGT29560.1
Cry3Ba1	CAA34983.1	Cry9Eb1	CAC50780.1	Cry32Ma1	AGU13855.1	Cry65Aa1	AEB52307.1
Cry3Bb1	AAA22334.1	Cry9Ec1	AAC63366.2	Cry32Mb1	AGU13869.1	Cry66Aa2	ZP_04099945.1

Holotyp	RefSeq/GenBank Sequenz						
Cry3Ca1	CAA42469.1	Cry9Ed1	AAX78440.1	Cry32Na1	AGU13856.1	Cry67Aa2	ZP_04148882.1
Cry4Aa1	CAA68485.1	Cry9Ee1	ADE60738.1	Cry32Oa1	AGU13868.1	Cry68Aa1	ADV33304.1
Cry4Ba1	CAA30312.1	Cry9Fa1	AGU13857.1	Cry32Pa1	AGU13870.1	Cry69Aa1	ADV77221.1
Cry4Ca1	ACF60501.1	Cry9Ga1	AGU13864.1	Cry32Qa1	AGU13871.1	Cry69Ab1	AER28313.1
Cry4Cb1	ACQ91257.1	Cry10Aa1	AAA22614.1	Cry32Ra1	AGU13872.1	Cry70Aa1	AEX56524.1
Cry4Cc1	ACQ91256.1	Cry11Aa1	AAA22352.1	Cry32Sa1	AGU13874.1	Cry70Ba1	ADO51070.1
Cry5Aa1	AAA67694.1	Cry11Ba1	CAA60504.1	Cry32Ta1	AGU13875.1	Cry70Bb1	EEL67276.1
Cry5Ab1	AAA67693.1	Cry11Bb1	AAC97162.1	Cry32Ua1	AGU13820.1	Cry73Aa1	AEH76822.1
Cry5Ac1	I34543.1	Cry12Aa1	AAA22355.1	Cry32Va1	WP_029439076.1	Cry74Aa1	WP_029440438.1
<b>Cyt</b>		<b>Vip</b>		<b>Vip</b>		<b>Vip</b>	
Cyt1Aa1	CAA26943.1	Vip1Aa2	AAR81088.1	Vip2Af1	ACH42759.1	Vip3Ai1	AGU13858.1
Cyt1Ab1	CAA67328.1	Vip1Ac1	AEH05932.1	Vip2Ag1	AGC08396.1	Vip3Aj1	AIT93172.1
Cyt1Ba1	AAB03693.1	Vip1Ad1	AGC08395.1	Vip2Ba1	AAR40887.1	Vip3Ba1	AAV70653.1
Cyt1Ca1	CAD30104.1	Vip1Ba1	AAR40886.1	Vip2Bb3	AIA96500.1	Vip3Bb2	ABO30520.1
Cyt1Da1	ADV33305.1	Vip1Bb1	AAR40282.1	Vip3Aa1	AAC37036.1	Vip3Ca1	ADZ46178.1
Cyt2Aa1	CAA78519.1	Vip1Ca1	AAO86514.1	Vip3Ab1	AAR40284.1	Vip4Aa1	AEB52299.1
Cyt2Ba1	AAB63254.1	Vip1Da1	CAI40767.1	Vip3Ad2	CAI43276.1		
Cyt2Bb1	AAB93477.1	Vip2Aa1	1QS1_A	Vip3Ae1	CAI43277.1	<b>Sip</b>	
Cyt2Bc1	CAC80987.1	Vip2Ac1	AAO86513.1	Vip3Af1	CAI43275.1	Sip1Aa1	ABC71340.1
Cyt2Ca1	AAK50455.1	Vip2Ad1	CAI40768.1	Vip3Ag2	ACL97352.2		
Cyt3Aa1	ADM53423.1	Vip2Ae1	ABR68092.1	Vip3Ah1	ABH10614.1		

Tabelle 35 Angewendete RB:BR Volumenverhältnisse während der ersten Längenselektion

Schritt	Handlung	Zusätzliche Information	Kategorie der Sequenzierbibliothek			
			IS1	IS2	IS3	IS4
1	Verwendung von 50 µl gescherter DNA	DNA ist gelöst in RB	50,0	50,0	50,0	50,0
2	Zugabe von RB		+ 30,0	+ 38,0	+ 42,0	+ 46,0
RB Volumen nach Schritt 2			<b>80,0</b>	<b>88,0</b>	<b>92,0</b>	<b>96,0</b>
3	Zugabe von 1X BR		+ 40,0	+ 40,0	+ 40,0	+ 40,0
Gesamtvolumen (RB und BR) nach Schritt 3			<b>120,0</b>	<b>128,0</b>	<b>132,0</b>	<b>136,0</b>
RB:BR Volumenverhältnis			80,0:40,0 (2,0:1,0)	88,0:40,0 (2,2:1,0)	92,0:40,0 (2,3:1,0)	96,0:40,0 (2,4:1,0)
4	Entnahme des Überstandes		117,0	125,0	129,0	133,0
5	Zugabe von RB		+ 70,2	+ 78,3	+ 82,1	+ 86,0
6	Zugabe von 2X BR		+ 39,0	+ 39,1	+ 39,1	+ 39,1
Gesamtvolumen (RB und BR) nach Schritt 6			<b>226,2</b>	<b>242,4</b>	<b>250,2</b>	<b>258,1</b>
RB:BB ratio by volume			148,2:78,0 (~ 3,8:2,0)	164,2:78,2 (~ 4,2:2,0)	172,0:78,2 (~ 4,4:2,0)	179,9:78,2 (~ 4,6:2,0)
7	Fahre fort mit Bead-Pellet wie in M-Abschnitt 6.1.2. für IS1-Bibliotheken beschrieben		...	...	...	...

**Farbschema**

Blau – Volumen RB in µl

Rot – Volumen BR in µl, wobei der BR Puffer 1X oder 2X konzentrierte magnetische Beads enthält

Magenta – Volumen von RB und BR Gemischen in µl

Grau schattierte Zeilen – Informationen über die Volumenzusammensetzung von RB und BR Gemischen in µl, sowie zugrundeliegende Volumenverhältnisse in schwarzen Klammern

Diese Tabelle wurde ins Deutsche übersetzt und entspricht Tabelle S1 aus der Studie von Huptas *et al.* [342].Die Tabelle unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Tabelle 36 Zugrundeliegende Daten der Logit-Modelle

Genompaar <i>i</i>	Stamm 1	Stamm 2	DDH	$Y = y_i$	$X = x_i$		
				(DDH $\geq$ 70 % ?)	ANib <sub>11</sub>	ANib <sub>12</sub>	ANIm
<b>Bacillus</b>							
1	<i>B. cereus</i> ATCC 10987	<i>B. cereus</i> ATCC 14579 (T)	63,80 <sup>a,∅</sup>	0	91,36	92,12	91,97
<b>Bacteroides</b>							
2	<i>B. cellulosilyticus</i> DSM 14838 (T)	<i>B. intestinalis</i> DSM 17393 (T)	40,00 <sup>c</sup>	0	86,98	88,62	90,25
3	<i>B. dorei</i> DSM 17855 (T)	<i>B. vulgatus</i> ATCC 8482 (T)	46,00 <sup>c</sup>	0	94,40	95,76	95,84
4	<i>B. finegoldii</i> DSM 17565 (T)	<i>B. ovatus</i> ATCC 8483 (T)	23,00 <sup>c</sup>	0	83,56	85,61	88,02
5	<i>B. finegoldii</i> DSM 17565 (T)	<i>B. thetaiotaomicron</i> VPI-5482 (T)	20,00 <sup>c</sup>	0	80,05	83,48	85,89
6	<i>B. fragilis</i> NCTC 9343 (T)	<i>B. fragilis</i> YCH46	99,00 <sup>b</sup>	1	98,59	99,00	99,17
7	<i>B. fragilis</i> NCTC 9343 (T)	<i>B. thetaiotaomicron</i> VPI-5482 (T)	24,50 <sup>b,∅</sup>	0	74,13	77,78	84,29
8	<i>B. intestinalis</i> DSM 17393 (T)	<i>B. uniformis</i> ATCC 8492 (T)	9,00 <sup>c</sup>	0	75,44	78,91	84,06
<b>Bifidobacterium</b>							
9	<i>B. angulatum</i> DSM 20098 (T)	<i>B. gallicum</i> DSM 20093 (T)	18,00 <sup>c</sup>	0	72,65	77,28	84,07
10	<i>B. breve</i> DSM 20213 (T)	<i>B. gallicum</i> DSM 20093 (T)	11,00 <sup>c</sup>	0	72,59	77,99	84,04
11	<i>B. catenulatum</i> DSM 16992 (T)	<i>B. gallicum</i> DSM 20093 (T)	15,00 <sup>c</sup>	0	72,89	78,73	84,82
12	<i>B. gallicum</i> DSM 20093 (T)	<i>B. pseudocatenulatum</i> DSM 20438 (T)	21,00 <sup>c</sup>	0	73,02	77,65	84,54
<b>Brucella</b>							
13	<i>B. canis</i> ATCC 23365 (T)	<i>B. melitensis</i> ATCC 23457	98,00 <sup>b,∅</sup>	1	99,67	99,76	99,64
14	<i>B. canis</i> ATCC 23365 (T)	<i>B. ovis</i> ATCC 25840 (T)	97,00 <sup>b,∅</sup>	1	99,59	99,75	99,57
15	<i>B. canis</i> ATCC 23365 (T)	<i>B. suis</i> 1330 (T)	100,00 <sup>b,∅</sup>	1	99,90	99,93	99,87
16	<i>B. melitensis</i> ATCC 23457	<i>B. ovis</i> ATCC 25840 (T)	95,50 <sup>b,∅</sup>	1	99,54	99,72	99,54
17	<i>B. melitensis</i> ATCC 23457	<i>B. suis</i> 1330 (T)	96,00 <sup>b,∅</sup>	1	99,70	99,78	99,66
18	<i>B. ovis</i> ATCC 25840 (T)	<i>B. suis</i> 1330 (T)	96,50 <sup>b,∅</sup>	1	99,63	99,77	99,60
19	<i>B. suis</i> ATCC 23445	<i>B. melitensis</i> ATCC 23457	96,00 <sup>b</sup>	1	99,65	99,75	99,60
<b>Burkholderia</b>							
20	<i>B. cenocepacia</i> AU 1054	<i>B. vietnamiensis</i> G4	31,40 <sup>a,∅</sup>	0	87,95	90,92	90,26

Genompaar <i>i</i>	Stamm 1	Stamm 2	DDH	$Y = y_i$	$X = x_i$		
				(DDH $\geq$ 70 % ?)	ANib <sub>11</sub>	ANib <sub>12</sub>	ANIm
21	<i>B. cenocepacia</i> AU 1054	<i>B. xenovorans</i> LB400 (T)	13,30 <sup>a,0</sup>	0	77,00	82,15	84,82
22	<i>B. cenocepacia</i> HI2424	<i>B. cenocepacia</i> J2315 (T)	67,45 <sup>a,0</sup>	0	94,53	96,01	95,09
23	<i>B. cenocepacia</i> HI2424	<i>B. lata</i> 383 (T)	51,40 <sup>a,0</sup>	0	90,37	93,02	91,73
24	<i>B. cenocepacia</i> HI2424	<i>B. vietnamiensis</i> G4	31,95 <sup>a,0</sup>	0	87,98	91,12	90,24
25	<i>B. cenocepacia</i> HI2424	<i>B. xenovorans</i> LB400 (T)	12,60 <sup>a,0</sup>	0	76,96	82,28	84,82
26	<i>B. cenocepacia</i> J2315 (T)	<i>B. lata</i> 383 (T)	55,15 <sup>a,0</sup>	0	90,57	93,07	91,96
27	<i>B. cenocepacia</i> J2315 (T)	<i>B. vietnamiensis</i> G4	35,60 <sup>a,0</sup>	0	87,95	91,28	90,41
28	<i>B. cenocepacia</i> J2315 (T)	<i>B. xenovorans</i> LB400 (T)	13,35 <sup>a,0</sup>	0	76,80	82,18	84,81
29	<i>B. lata</i> 383 (T)	<i>B. vietnamiensis</i> G4	31,15 <sup>a,0</sup>	0	87,08	90,73	89,81
30	<i>B. lata</i> 383 (T)	<i>B. xenovorans</i> LB400 (T)	13,20 <sup>a,0</sup>	0	76,52	82,01	84,74
31	<i>B. vietnamiensis</i> G4	<i>B. xenovorans</i> LB400 (T)	9,55 <sup>a,0</sup>	0	76,86	82,23	84,89
<b>Chloroflexus</b>							
32	<i>C. aggregans</i> DSM 9485 (T)	<i>C. aurantiacus</i> J-10-fl (T)	10,00 <sup>b</sup>	0	74,10	76,75	82,88
<b>Clostridium</b>							
33	<i>C. hylemonae</i> DSM 15053 (T)	<i>C. scindens</i> ATCC 35704 (T)	11,50 <sup>c</sup>	0	73,60	76,73	83,49
<b>Corynebacterium</b>							
34	<i>C. efficiens</i> YS-314 (T)	<i>C. glutamicum</i> ATCC 13032 (T)	8,00 <sup>c</sup>	0	73,64	77,08	84,64
35	<i>C. striatum</i> ATCC 6940 (T)	<i>C. urealyticum</i> DSM 7109 (T)	6,00 <sup>c</sup>	0	70,44	76,61	83,88
<b>Cupriavidus/Ralstonia</b>							
36	<i>C. metallidurans</i> CH34 (T)	<i>R. eutropha</i> H16	28,00 <sup>b</sup>	0	78,90	83,52	85,73
<b>Deinococcus</b>							
37	<i>D. deserti</i> VCD115 (T)	<i>D. radiodurans</i> R1 (T)	17,00 <sup>b</sup>	0	74,08	78,07	83,99
<b>Escherichia/Shigella</b>							
38	<i>E. coli</i> 042	<i>E. coli</i> CFT073	82,95 <sup>a,0</sup>	1	96,46	96,97	96,97
39	<i>E. coli</i> 042	<i>E. coli</i> K-12 MG1655	73,60 <sup>a,0</sup>	1	97,27	97,59	97,58

Genompaar <i>i</i>	Stamm 1	Stamm 2	DDH	$Y = y_i$	$X = x_i$		
				(DDH $\geq$ 70 % ?)	ANib <sub>11</sub>	ANib <sub>12</sub>	ANIm
40	<i>E. coli</i> 042	<i>E. coli</i> O127:H6 E2348/69	75,05 <sup>a,0</sup>	1	96,61	97,08	97,09
41	<i>E. coli</i> 042	<i>E. coli</i> O157:H7 EDL933	75,00 <sup>a,0</sup>	1	96,52	96,99	97,12
42	<i>E. coli</i> 042	<i>E. coli</i> O157:H7 Sakai	77,65 <sup>a,0</sup>	1	96,54	97,02	97,06
43	<i>E. coli</i> 042	<i>S. flexneri</i> 2a 2457 (T)	66,80 <sup>a,0</sup>	0	97,03	97,41	97,37
44	<i>E. coli</i> 042	<i>S. sonnei</i> 53G	71,40 <sup>a,0</sup>	1	96,94	97,39	97,45
45	<i>E. coli</i> CFT073	<i>E. coli</i> K-12 MG1655	92,35 <sup>a,0</sup>	1	96,72	97,11	97,15
46	<i>E. coli</i> CFT073	<i>E. coli</i> O127:H6 E2348/69	76,45 <sup>a,0</sup>	1	98,21	98,54	98,49
47	<i>E. coli</i> CFT073	<i>E. coli</i> O157:H7 EDL933	80,60 <sup>a,0</sup>	1	96,43	96,92	96,92
48	<i>E. coli</i> CFT073	<i>E. coli</i> O157:H7 Sakai	84,65 <sup>a,0</sup>	1	96,43	96,95	96,87
49	<i>E. coli</i> CFT073	<i>S. flexneri</i> 2a 2457 (T)	72,50 <sup>a,0</sup>	1	96,64	97,07	97,02
50	<i>E. coli</i> CFT073	<i>S. sonnei</i> 53G	78,05 <sup>a,0</sup>	1	96,55	97,08	97,12
51	<i>E. coli</i> K-12 MG1655	<i>E. coli</i> O127:H6 E2348/69	75,60 <sup>a,0</sup>	1	96,90	97,21	97,19
52	<i>E. coli</i> K-12 MG1655	<i>E. coli</i> O157:H7 EDL933	79,25 <sup>a,0</sup>	1	97,86	98,12	98,06
53	<i>E. coli</i> K-12 MG1655	<i>E. coli</i> O157:H7 Sakai	80,35 <sup>a,0</sup>	1	97,87	98,15	98,07
54	<i>E. coli</i> K-12 MG1655	<i>S. flexneri</i> 2a 2457 (T)	68,10 <sup>a,0</sup>	0	98,17	98,37	98,31
55	<i>E. coli</i> K-12 MG1655	<i>S. sonnei</i> 53G	74,80 <sup>a,0</sup>	1	98,33	98,53	98,48
56	<i>E. coli</i> O127:H6 E2348/69	<i>E. coli</i> O157:H7 EDL933	76,20 <sup>a,0</sup>	1	96,50	96,97	96,93
57	<i>E. coli</i> O127:H6 E2348/69	<i>E. coli</i> O157:H7 Sakai	79,45 <sup>a,0</sup>	1	96,47	97,00	96,94
58	<i>E. coli</i> O127:H6 E2348/69	<i>S. flexneri</i> 2a 2457 (T)	66,25 <sup>a,0</sup>	0	96,68	97,09	97,04
59	<i>E. coli</i> O127:H6 E2348/69	<i>S. sonnei</i> 53G	70,95 <sup>a,0</sup>	1	96,67	97,10	97,12
60	<i>E. coli</i> O157:H7 EDL933	<i>E. coli</i> O157:H7 Sakai	98,85 <sup>a,0</sup>	1	99,91	99,92	99,57
61	<i>E. coli</i> O157:H7 EDL933	<i>S. flexneri</i> 2a 2457 (T)	67,80 <sup>a,0</sup>	0	97,46	97,81	97,81
62	<i>E. coli</i> O157:H7 EDL933	<i>S. sonnei</i> 53G	74,35 <sup>a,0</sup>	1	97,39	97,83	97,76
63	<i>E. coli</i> O157:H7 Sakai	<i>S. flexneri</i> 2a 2457 (T)	65,90 <sup>a,0</sup>	0	97,43	97,82	97,83
64	<i>E. coli</i> O157:H7 Sakai	<i>S. sonnei</i> 53G	74,00 <sup>a,0</sup>	1	97,30	97,78	97,77
65	<i>S. flexneri</i> 2a 2457 (T)	<i>S. sonnei</i> 53G	79,70 <sup>a,0</sup>	1	98,30	98,59	98,53
<b>Halorubrum/Haloterrigena</b>							
66	<i>H. lacusprofundi</i> ATCC 49239 (T)	<i>H. turkmenica</i> DSM 5511 (T)	28,00 <sup>c</sup>	0	72,46	77,96	82,93

**Methanococcus**

Genompaar <i>i</i>	Stamm 1	Stamm 2	DDH	$Y = y_i$	$X = x_i$		
				(DDH $\geq$ 70 % ?)	ANib <sub>11</sub>	ANib <sub>12</sub>	ANIm
67	<i>M. maripaludis</i> C5	<i>M. maripaludis</i> C6	69,00 <sup>b</sup>	0	90,83	91,62	91,70
68	<i>M. maripaludis</i> C5	<i>M. maripaludis</i> C7	69,00 <sup>b</sup>	0	89,70	90,54	90,58
69	<i>M. maripaludis</i> C5	<i>M. vannielii</i> SB (T)	17,00 <sup>b</sup>	0	76,71	79,51	83,16
70	<i>M. maripaludis</i> C6	<i>M. maripaludis</i> C7	81,00 <sup>b</sup>	1	90,53	91,12	90,93
71	<i>M. maripaludis</i> C6	<i>M. vannielii</i> SB (T)	27,00 <sup>b</sup>	0	76,74	79,46	83,26
72	<i>M. maripaludis</i> C7	<i>M. vannielii</i> SB (T)	32,00 <sup>b</sup>	0	76,72	79,57	83,71
<b>Nitrosomonas</b>							
73	<i>N. europaea</i> ATCC 19718	<i>N. eutropha</i> C91	43,00 <sup>b</sup>	0	79,01	80,51	83,20
<b>Parabacteroides</b>							
74	<i>P. distasonis</i> ATCC 8503 (T)	<i>P. johnsonii</i> DSM 18315 (T)	13,00 <sup>c</sup>	0	73,50	76,71	84,00
75	<i>P. distasonis</i> ATCC 8503 (T)	<i>P. merdae</i> ATCC 43184 (T)	15,00 <sup>c</sup>	0	74,02	76,90	86,64
76	<i>P. johnsonii</i> DSM 18315 (T)	<i>P. merdae</i> ATCC 43184 (T)	59,00 <sup>c</sup>	0	90,33	91,39	91,87
<b>Pseudomonas</b>							
77	<i>P. aeruginosa</i> PAO1	<i>P. fluorescens</i> Pf0-1	13,45 <sup>a,∅</sup>	0	75,38	79,30	83,54
78	<i>P. aeruginosa</i> PAO1	<i>P. fluorescens</i> SBW25	12,60 <sup>a,∅</sup>	0	74,83	78,85	83,49
79	<i>P. aeruginosa</i> PAO1	<i>P. protegens</i> Pf-5	14,00 <sup>a,∅</sup>	0	76,09	80,02	83,84
80	<i>P. aeruginosa</i> PAO1	<i>P. syringae</i> B728a	8,05 <sup>a,∅</sup>	0	74,37	78,57	83,25
81	<i>P. aeruginosa</i> PAO1	<i>P. syringae</i> DC3000	6,20 <sup>a,∅</sup>	0	74,08	78,15	83,20
82	<i>P. fluorescens</i> Pf0-1	<i>P. fluorescens</i> SBW25	29,35 <sup>a,∅</sup>	0	80,66	83,17	85,83
83	<i>P. fluorescens</i> Pf0-1	<i>P. protegens</i> Pf-5	27,60 <sup>a,∅</sup>	0	81,09	83,60	85,73
84	<i>P. fluorescens</i> Pf0-1	<i>P. syringae</i> B728a	17,90 <sup>a,∅</sup>	0	77,44	80,71	84,52
85	<i>P. fluorescens</i> Pf0-1	<i>P. syringae</i> DC3000	14,15 <sup>a,∅</sup>	0	77,44	80,68	84,40
86	<i>P. fluorescens</i> SBW25	<i>P. protegens</i> Pf-5	27,25 <sup>a,∅</sup>	0	80,49	83,09	85,51
87	<i>P. fluorescens</i> SBW25	<i>P. syringae</i> B728a	16,40 <sup>a,∅</sup>	0	77,06	80,25	84,53
88	<i>P. fluorescens</i> SBW25	<i>P. syringae</i> DC3000	13,70 <sup>a,∅</sup>	0	76,97	80,17	84,39
89	<i>P. protegens</i> Pf-5	<i>P. syringae</i> B728a	13,40 <sup>a,∅</sup>	0	77,15	80,60	84,68
90	<i>P. protegens</i> Pf-5	<i>P. syringae</i> DC3000	10,85 <sup>a,∅</sup>	0	77,10	80,51	84,45
91	<i>P. syringae</i> B728a	<i>P. syringae</i> DC3000	38,15 <sup>a,∅</sup>	0	86,07	87,22	87,97

Genompaar <i>i</i>	Stamm 1	Stamm 2	DDH	$Y = y_i$	$X = x_i$		
				(DDH $\geq$ 70 % ?)	ANib <sub>11</sub>	ANib <sub>12</sub>	ANIm
<b><i>Pyrobaculum</i></b>							
92	<i>P. aerophilum</i> IM2 (T)	<i>P. calidifontis</i> JCM 11548 (T)	18,00 <sup>b</sup>	0	69,46	73,46	86,93
93	<i>P. aerophilum</i> IM2 (T)	<i>P. islandicum</i> DSM 4184 (T)	3,00 <sup>b</sup>	0	69,49	72,87	93,44
94	<i>P. calidifontis</i> JCM 11548 (T)	<i>P. islandicum</i> DSM 4184 (T)	1,00 <sup>b</sup>	0	68,54	72,63	86,97
<b><i>Salinibacter</i></b>							
95	<i>S. ruber</i> DSM 13855 (T)	<i>S. ruber</i> M8	96,60 <sup>b</sup>	1	98,51	98,78	98,65
<b><i>Selenomonas</i></b>							
96	<i>S. noxia</i> ATCC 43541 (T)	<i>S. sputigena</i> ATCC 35185 (T)	13,00 <sup>c</sup>	0	69,34	75,86	83,25
<b><i>Shewanella</i></b>							
97	<i>S. amazonensis</i> SB2B (T)	<i>S. oneidensis</i> MR-1 (T)	40,30 <sup>b</sup>	0	70,83	75,53	83,09
98	<i>S. amazonensis</i> SB2B (T)	<i>S. pealeana</i> ATCC 700345 (T)	27,30 <sup>c</sup>	0	69,55	74,58	83,56
99	<i>S. amazonensis</i> SB2B (T)	<i>S. putrefaciens</i> CN-32	42,70 <sup>b</sup>	0	70,49	75,06	82,21
100	<i>S. amazonensis</i> SB2B (T)	<i>S. woodyi</i> ATCC 51908 (T)	40,00 <sup>b</sup>	0	69,18	74,31	82,98
101	<i>S. baltica</i> OS155	<i>S. baltica</i> OS185	99,00 <sup>b</sup>	1	96,65	97,07	96,88
102	<i>S. baltica</i> OS155	<i>S. baltica</i> OS195	99,00 <sup>b</sup>	1	96,58	97,02	96,80
103	<i>S. baltica</i> OS155	<i>S. baltica</i> OS223	99,00 <sup>b</sup>	1	96,86	97,25	97,09
104	<i>S. baltica</i> OS155	<i>S. oneidensis</i> MR-1 (T)	25,95 <sup>a,∅</sup>	0	79,21	81,15	84,63
105	<i>S. baltica</i> OS155	<i>S. putrefaciens</i> CN-32	26,95 <sup>a,∅</sup>	0	82,70	84,25	86,67
106	<i>S. halifaxensis</i> HAW-EB4 (T)	<i>S. pealeana</i> ATCC 700345 (T)	17,90 <sup>b</sup>	0	82,22	83,77	85,89
107	<i>S. oneidensis</i> MR-1 (T)	<i>S. pealeana</i> ATCC 700345 (T)	20,40 <sup>b</sup>	0	71,43	75,90	83,22
108	<i>S. oneidensis</i> MR-1 (T)	<i>S. putrefaciens</i> CN-32	21,60 <sup>a,∅</sup>	0	79,27	81,29	85,01
109	<i>S. oneidensis</i> MR-1 (T)	<i>S. woodyi</i> ATCC 51908 (T)	39,00 <sup>b</sup>	0	71,06	75,44	83,06
110	<i>S. sediminis</i> HAW-EB3 (T)	<i>S. woodyi</i> ATCC 51908 (T)	17,80 <sup>b</sup>	0	75,27	78,31	84,42
<b><i>Streptococcus</i></b>							
111	<i>S. agalactiae</i> 2603V/R	<i>S. agalactiae</i> A909	93,45 <sup>a,∅</sup>	1	99,25	99,44	99,43
112	<i>S. agalactiae</i> 2603V/R	<i>S. agalactiae</i> NEM316	86,35 <sup>a,∅</sup>	1	98,73	99,02	99,42

Genompaar <i>i</i>	Stamm 1	Stamm 2	DDH	$Y = y_i$	$X = x_i$		
				(DDH $\geq$ 70 % ?)	ANib <sub>11</sub>	ANib <sub>12</sub>	ANIm
113	<i>S. agalactiae</i> A909	<i>S. agalactiae</i> NEM316	90,65 <sup>a,Ø</sup>	1	99,17	99,33	99,34
<b><i>Sulfolobus</i></b>							
114	<i>S. acidocaldarius</i> DSM 639 (T)	<i>S. solfataricus</i> P2	4,55 <sup>b,Ø</sup>	0	66,37	72,31	89,22
<b><i>Thermotoga</i></b>							
115	<i>T. maritima</i> MSB8 (T)	<i>T. neapolitana</i> DSM 4359 (T)	30,80 <sup>b</sup>	0	79,99	81,39	87,17
116	<i>T. neapolitana</i> DSM 4359 (T)	<i>T. petrophila</i> RKU-1 (T)	11,00 <sup>b</sup>	0	79,04	80,53	85,77
<b><i>Yersinia</i></b>							
117	<i>Y. aldovae</i> ATCC 35236 (T)	<i>Y. bercovieri</i> ATCC 43970 (T)	38,00 <sup>c</sup>	0	82,47	82,34	85,64
118	<i>Y. aldovae</i> ATCC 35236 (T)	<i>Y. mollaretii</i> ATCC 43969 (T)	32,00 <sup>c</sup>	0	82,33	82,90	85,53
119	<i>Y. bercovieri</i> ATCC 43970 (T)	<i>Y. frederiksenii</i> ATCC 33641 (T)	44,00 <sup>c</sup>	0	83,12	84,26	85,88
120	<i>Y. bercovieri</i> ATCC 43970 (T)	<i>Y. intermedia</i> ATCC 29909 (T)	40,00 <sup>c</sup>	0	83,79	84,86	86,18
121	<i>Y. bercovieri</i> ATCC 43970 (T)	<i>Y. mollaretii</i> ATCC 43969 (T)	59,00 <sup>c</sup>	0	89,13	89,60	90,03
122	<i>Y. bercovieri</i> ATCC 43970 (T)	<i>Y. rohdei</i> ATCC 43380 (T)	40,00 <sup>c</sup>	0	82,92	83,42	85,67
123	<i>Y. frederiksenii</i> ATCC 33641 (T)	<i>Y. mollaretii</i> ATCC 43969 (T)	42,00 <sup>c</sup>	0	83,17	84,85	85,79
124	<i>Y. frederiksenii</i> ATCC 33641 (T)	<i>Y. rohdei</i> ATCC 43380 (T)	53,00 <sup>c</sup>	0	85,07	85,92	86,79
125	<i>Y. intermedia</i> ATCC 29909 (T)	<i>Y. mollaretii</i> ATCC 43969 (T)	33,00 <sup>c</sup>	0	83,72	85,26	86,04
126	<i>Y. intermedia</i> ATCC 29909 (T)	<i>Y. rohdei</i> ATCC 43380 (T)	45,00 <sup>c</sup>	0	82,73	84,21	85,59
127	<i>Y. mollaretii</i> ATCC 43969 (T)	<i>Y. rohdei</i> ATCC 43380 (T)	22,00 <sup>c</sup>	0	82,96	83,96	85,60

<sup>a</sup>, DDH Daten von Goris *et al.* [47].

<sup>b</sup>, DDH Daten von Richter *et al.* [48].

<sup>c</sup>, DDH Daten aus der GOLD Datenbank [386].

<sup>Ø</sup>, Durchschnitt von DDH(Stamm 1, Stamm 2) und DDH(Stamm 2, Stamm 1)

Tabelle 37 Durchschnittliche Insert-Länge sequenzierter Bibliotheken

Sequenzierte Bibliothek	Durchschnittliche Insert-Länge in bp basierend auf		Überschätzung Bioanalyser in %
	Bioanalyser Messung	Rohdaten Remapping*	
Efa_IS1	704	674	4,45
Mlu_IS1	719	675	6,52
Bce_IS1	732	669	9,42
Pst_IS1	819	714	14,71
Sen_IS1	886	726	22,04
Efa_IS2	1.105	918	20,37
Sen_IS2	1.376	978	40,70
Mlu <sub>50</sub> _IS2	1.408	1.074	31,10
Bce_IS3	1.602	1.094	46,44
Mlu_IS3	1.707	1.255	36,02
Mlu <sub>50</sub> _IS3	1.709	1.246	37,16
Pst_IS3	1.734	1.250	38,72

\* Werte inkl. 20 bp aufgrund des Trimmens der 5'-Read-Enden (Methoden-Abschnitt 6.4.)

Diese Tabelle wurde ins Deutsche übersetzt und entspricht Tabelle S3 aus der Studie von Huptas *et al.* [342].

Die Tabelle unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Tabelle 38 Durchschnittliche Insert-Längen von Bibliotheken vor und nach der Qualitätskontrolle

Sequenzierte Bibliothek	% hoch-qualitative Read-Paare**	Durchschnittliche Insert-Länge* in bp basierend auf		% Verringerung der durchschnittlichen Insert-Länge
		Remapping ungefilterte Bibliothek	Remapping gefilterte Bibliothek**	
Bce_TS	85,66	654	632	3,36
Bce_IS1	86,45	669	661	1,20
Bce_IS3	71,43	1.094	1.038	5,12
Efa_TS	94,02	610	597	2,13
Efa_IS1	93,72	674	668	0,89
Efa_IS2	69,81	918	877	4,47
Sen_IS1	80,74	726	713	1,79
Sen_IS2	50,09	978	915	6,44
Pst_TS	75,50	673	647	3,86
Pst_IS1	71,22	714	698	2,24
Pst_IS3	40,10	1.250	1.121	10,32
Mlu_TS	75,42	626	604	3,51
Mlu_IS1	70,89	675	664	1,63
Mlu_IS3	34,81	1.255	1.107	11,79

\* Werte inkl. 20 bp aufgrund des Trimmens der 5'-Read-Enden (Methoden-Abschnitt 6.4. )

\*\* Hoch-qualitative Read-Paare weisen einen Phred Qualitätswert  $\geq 20$  über mindestens 80 % der Read-Länge auf (Methoden-Abschnitt 6.4. )

Diese Tabelle wurde ins Deutsche übersetzt und entspricht Tabelle S4 aus der Studie von Huptas *et al.* [342].

Die Tabelle unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Tabelle 39 Relative Assemblierungs-Werte untersuchter durchschnittlicher Insert-Längen (ABySS)

Assemblierungs-Metrik	Ø Insert-Länge	Genom			
		Bce	Efa	Sen	Pst
korrigierter NG50	TS	0,81	1	n.d.	0,74
	IS1	0,82	0,99	0,96	0,69
	IS2 oder IS3	1	0,98	1	1
	<i>max</i>	353.214	335.452	246.064	183.457
NGA50	TS	0,86	0,85	n.d.	0,98
	IS1	0,83	0,87	0,77	0,95
	IS2 oder IS3	1	1	1	1
	<i>max</i>	759.632	445.272	458.061	223.858

*max* Absoluter Wert in Nukleotiden des maximalen relativen Assemblierungs-Wertes der Spalte.

Diese Tabelle wurde ins Deutsche übersetzt und entspricht Tabelle S6 aus der Studie von Huptas *et al.* [342].

Die Tabelle unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Tabelle 40 Relative Assemblierungs-Werte untersuchter Read-Längen (ABySS)

Assemblierungs-Metrik	Read-Länge [bp]	Genom			
		Bce	Efa	Sen	Pst
korrigierter NG50	100	0,89	0,78	0,85	1
	125	0,94	0,82	0,999	0,80
	150	0,72	0,91	0,94	0,75
	175	1	0,99	1	0,90
	189	0,75	1	0,998	0,86
	<i>max</i>	360.330	368.221	251.266	172.209
NGA50	100	0,74	0,96	0,80	0,98
	125	0,75	0,96	1	0,99
	150	1	0,94	0,98	0,97
	175	0,94	1	0,92	0,99
	189	0,90	0,96	0,90	1
	<i>max</i>	784.943	418.320	440.832	221.474

*max* Absoluter Wert in Nukleotiden des maximalen relativen Assemblierungs-Wertes der Spalte.

Diese Tabelle wurde ins Deutsche übersetzt und entspricht Tabelle S7 aus der Studie von Huptas *et al.* [342].

Die Tabelle unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Tabelle 41 Relative Assemblierungs-Werte untersuchter Sequenzertiefen (ABySS)

Assemblierungs-Metrik	Tiefe der Sequenzierung [x-fach]	Genom		
		Bce	Efa	Pst
korrigierter NG50	45	0,74	0,70	0,77
	90	1	1	1
	<i>max</i>	355.983	389.866	167.776
NGA50	45	0,76	0,90	0,96
	90	1	1	1
	<i>max</i>	775.233	424.126	223.188

*max* Absoluter Wert in Nukleotiden des maximalen relativen Assemblierungs-Wertes der Spalte.

Diese Tabelle wurde ins Deutsche übersetzt und entspricht Tabelle S8 aus der Studie von Huptas *et al.* [342].

Die Tabelle unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Tabelle 42 Relative Assemblierungs-Werte untersuchter durchschnittlicher Insert-Längen (Velvet)

Assemblierungs-Metrik	Ø Insert-Länge	Genom			
		Bce	Efa	Sen	Pst
korrigierter NG50	TS	0,34	1	n.d.	0,94
	IS1	1	0,88	1	0,96
	IS2 oder IS3	0,35	0,82	0,94	1
	<i>max</i>	567.563	300.545	231.909	206.284
NGA50	TS	0,30	1	n.d.	0,97
	IS1	1	0,85	1	0,97
	IS2 oder IS3	0,37	0,87	0,92	1
	<i>max</i>	757.995	328.125	390.653	232.915

*max* Absoluter Wert in Nukleotiden des maximalen relativen Assemblierungs-Wertes der Spalte.

Diese Tabelle wurde ins Deutsche übersetzt und entspricht Tabelle S9 aus der Studie von Huptas *et al.* [342].

Die Tabelle unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Tabelle 43 Relative Assemblierungs-Werte untersuchter Read-Längen (Velvet)

Assemblierungs- Metrik	Read-Länge [bp]	Genom			
		Bce	Efa	Sen	Pst
korrigierter NG50	100	0,66	0,69	0,80	1
	125	0,82	0,69	0,83	0,91
	150	0,96	0,85	0,83	0,94
	175	1	1	1	0,95
	189	0,98	0,95	0,86	0,89
	<i>max</i>		373.968	323.927	259.132
NGA50	100	0,78	0,81	1	1
	125	0,87	0,87	0,99	0,97
	150	1	0,84	0,95	0,96
	175	0,89	1	0,81	0,91
	189	0,75	0,99	0,84	0,93
	<i>max</i>		490.929	379.037	408.297

*max* Absoluter Wert in Nukleotiden des maximalen relativen Assemblierungs-Wertes der Spalte.

Diese Tabelle wurde ins Deutsche übersetzt und entspricht Tabelle S10 aus der Studie von Huptas *et al.* [342].

Die Tabelle unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Tabelle 44 Relative Assemblierungs-Werte untersuchter Sequenziertiefen (Velvet)

Assemblierungs- Metrik	Tiefe der Sequenzierung [x-fach]	Genom		
		Bce	Efa	Pst
korrigierter NG50	45	0,70	0,59	0,87
	90	1	1	1
	<i>max</i>	375.042	341.912	213.284
NGA50	45	0,73	0,61	0,90
	90	1	1	1
	<i>max</i>	487.795	425.447	240.426

*max* Absoluter Wert in Nukleotiden des maximalen relativen Assemblierungs-Wertes der Spalte.

Diese Tabelle wurde ins Deutsche übersetzt und entspricht Tabelle S11 aus der Studie von Huptas *et al.* [342].

Die Tabelle unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Tabelle 45 Relative Assemblierungs-Werte untersuchter durchschnittlicher Insert-Längen (Edena)

Assemblierungs-Metrik	Ø Insert-Länge	Genom			
		Bce	Efa	Sen	Pst
korrigierter NG50	TS	0,20	1	n.d.	1
	IS1	1	0,99	1	0,96
	IS2 oder IS3	0,33	0,77	0,77	0,75
	<i>max</i>	551.178	296.040	209.880	171.430
NGA50	TS	0,21	1	n.d.	1
	IS1	1	0,998	1	0,98
	IS2 oder IS3	0,36	0,78	0,74	0,75
	<i>max</i>	560.150	297.278	234.630	171.623

*max* Absoluter Wert in Nukleotiden des maximalen relativen Assemblierungs-Wertes der Spalte.

Diese Tabelle wurde ins Deutsche übersetzt und entspricht Tabelle S12 aus der Studie von Huptas *et al.* [342].

Die Tabelle unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Tabelle 46 Relative Assemblierungs-Werte untersuchter Read-Längen (Edena)

Assemblierungs-Metrik	Read-Länge [bp]	Genom			
		Bce	Efa	Sen	Pst
korrigierter NG50	100	0,74	0,74	0,92	1
	125	0,99	0,83	0,98	0,99
	150	0,98	0,87	0,98	0,95
	175	0,98	1	1	0,81
	189	1	0,87	0,87	0,72
	<i>max</i>	299.749	317.133	195.108	173.950
NGA50	100	0,67	0,74	0,87	0,999
	125	0,91	0,83	0,95	1
	150	0,90	0,87	0,98	0,94
	175	1	1	1	0,81
	189	0,97	0,87	0,80	0,72
	<i>max</i>	329.643	318.278	221.295	174.717

*max* Absoluter Wert in Nukleotiden des maximalen relativen Assemblierungs-Wertes der Spalte.

Diese Tabelle wurde ins Deutsche übersetzt und entspricht Tabelle S13 aus der Studie von Huptas *et al.* [342].

Die Tabelle unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Tabelle 47 Relative Assemblierungs-Werte untersuchter Sequenzertiefen (Edena)

Assemblierungs- Metrik	Tiefe der Sequenzierung [x-fach]	Genom		
		Bce	Efa	Pst
korrigierter NG50	45	0,58	0,57	0,57
	90	1	1	1
	<i>max</i>	355.661	348.269	197.405
NGA50	45	0,57	0,58	0,57
	90	1	1	1
	<i>max</i>	374.324	348.271	198.796

*max* Absoluter Wert in Nukleotiden des maximalen relativen Assemblierungs-Wertes der Spalte.

Diese Tabelle wurde ins Deutsche übersetzt und entspricht Tabelle S14 aus der Studie von Huptas *et al.* [342].

Die Tabelle unterliegt der Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Tabelle 48 *Pseudomonas* Taxonomie nach LPSN mit Stand vom April 2017

Spezies nach LPSN		Referenz <sup>LPSN</sup>	Art des Typstammes nach LPSN		Typstamm im Datensatz (ID)
Nr.	Name		Spezies	Subspezies reklassifiziert	
1	<i>P. abietaniphila</i>	Mohn <i>et al.</i> 1999	x		<i>P. abietaniphila</i> ATCC 700689 (P101)
2	<i>P. aeruginosa</i>	Schroeter 1872	x		<i>P. aeruginosa</i> DSM 50071 (P001)
3	<i>P. aestusnigri</i>	Sanchez <i>et al.</i> 2014	x		<i>P. aestusnigri</i> CECT 8317 (P135)
4	<i>P. agarici</i>	Young 1970	x		<i>P. agarici</i> NCPPB 2289 (P136)
5	<i>P. alcaligenes</i>	Monias 1928	x		<i>P. alcaligenes</i> NBRC 14159 (P138)
6	<i>P. alcaliphila</i>	Yumoto <i>et al.</i> 2001	x		<i>P. alcaliphila</i> JCM 10630 (P140)
7	<i>P. alkylphenolica</i>	Mulet <i>et al.</i> 2015	x		<i>P. alkylphenolica</i> KL28 (P141)
8	<i>P. amygdali</i>	Psallidas & Panagopoulos 1975	x		<i>P. amygdali</i> ICMP 3918 (P142)
9	<i>P. anguilliseptica</i>	Wakabayashi & Egusa 1972	x		<i>P. anguilliseptica</i> DSM 12111 (P143)
10	<i>P. antarctica</i>	Reddy <i>et al.</i> 2004	x		<i>P. antarctica</i> LMG 22709 (P442) <sup>MLSA</sup>
11	<i>P. argentinensis</i>	Peix <i>et al.</i> 2005	x		<i>P. argentinensis</i> LMG 22563 (P144)
12	<i>P. arsenicoxydans</i>	Campos <i>et al.</i> 2011	x		<i>P. arsenicoxydans</i> CECT 7543 (P145)
13	<i>P. asplenii</i>	Ark & Tompkins 1946	x		<i>P. asplenii</i> ATCC 23835 (P146)
14	<i>P. asturiensis</i>	González <i>et al.</i> 2013,	x		<i>P. asturiensis</i> LMG 26898 (P147)
15	<i>P. asuensis</i>	Reddy & Garcia-Pichel 2015	x		---
16	<i>P. avellanae</i>	Janse <i>et al.</i> 1997	x		<i>P. avellanae</i> BPIC 631 (P148)
17	<i>P. azotifigens</i>	Hatayama <i>et al.</i> 2005	x		<i>P. azotifigens</i> DSM 17556 (P149)
18	<i>P. azotoformans</i>	Iizuka & Komagata 1963	x		<i>P. azotoformans</i> LMG 21611 (P150)
19	<i>P. baetica</i>	López <i>et al.</i> 2012	x		<i>P. baetica</i> CECT 7720 (P443) <sup>MLSA</sup>
20	<i>P. balearica</i>	Bennasar <i>et al.</i> 1996	x		<i>P. balearica</i> DSM 6083 (P151)
21	<i>P. bauzanensis</i>	Zhang <i>et al.</i> 2011	x		<i>P. bauzanensis</i> DSM 22558 (P152)
22	<i>P. benzenivorans</i>	Lang <i>et al.</i> 2012	x		<i>P. benzenivorans</i> DSM 8628 (P154)
23	<i>P. borbori</i>	Vanparys <i>et al.</i> 2006	x		<i>P. borbori</i> DSM 17834 (P155)
24	<i>P. brassicacearum</i> (ssp. <i>brassicacearum</i> )	Achouak <i>et al.</i> 2000	x	x	<i>P. brassicacearum</i> (ssp. <i>brassicacearum</i> ) CFBP 11706 (P444) <sup>MLSA</sup>
24	<i>P. brassicacearum</i> (ssp. <i>neoaurantiaca</i> )	Ivanova <i>et al.</i> 2009		x	---
25	<i>P. brenneri</i>	Baida <i>et al.</i> 2002	x		<i>P. brenneri</i> DSM 15294 (P004)
26	<i>P. caeni</i>	Xiao <i>et al.</i> 2009	x		<i>P. caeni</i> DSM 24390 (P160)
27	<i>P. cannabina</i>	ex Šutič & Dowson 1959	x		<i>P. cannabina</i> ICMP 2823 (P162)
28	<i>P. caricapapayae</i>	Robbs 1956	x		<i>P. caricapapayae</i> ICMP 2855 (P163)
29	<i>P. cedrina</i> (ssp. <i>cedrina</i> )	Dabboussi <i>et al.</i> 2002	x	x	<i>P. cedrina</i> (ssp. <i>cedrina</i> ) DSM 17516 (P164)
29	<i>P. cedrina</i> (ssp. <i>fulgida</i> )	Behrendt <i>et al.</i> 2009		x	<i>P. cedrina</i> (ssp. <i>fulgida</i> ) DSM 14938 (P445) <sup>MLSA</sup>
30	<i>P. chengduensis</i>	Tao <i>et al.</i> 2014	x		<i>P. chengduensis</i> DSM 26382 (P006)
31	<i>P. chlororaphis</i> (ssp. <i>aurantiaca</i> )	Nakhimovskaya 1948		x	<i>P. chlororaphis</i> (ssp. <i>aurantiaca</i> ) LMG 21630 (P167)
31	<i>P. chlororaphis</i> (ssp. <i>aureofaciens</i> )	Kluyver 1956		x	<i>P. chlororaphis</i> (ssp. <i>aureofaciens</i> ) NBRC 3521 (P168)
31	<i>P. chlororaphis</i> (ssp. <i>chlororaphis</i> )	Guignard & Sauvageau 1894	x	x	<i>P. chlororaphis</i> (ssp. <i>chlororaphis</i> ) DSM 50083 (P008)
31	<i>P. chlororaphis</i> (ssp. <i>piscium</i> )	Burr <i>et al.</i> 2010		x	<i>P. chlororaphis</i> (ssp. <i>piscium</i> ) DSM 21509 (P169)
32	<i>P. cichorii</i>	Swingle 1925	x		<i>P. cichorii</i> ATCC 10857 (P178)
33	<i>P. cissicola</i>	Takimoto 1939	x		---
34	<i>P. citronellolis</i>	Seubert 1960	x		<i>P. citronellolis</i> LMG 18378 (P180)
35	<i>P. coleopterorum</i>	Menéndez <i>et al.</i> 2015	x		<i>P. coleopterorum</i> LMG 28558 (P181)
36	<i>P. composti</i>	Gibello <i>et al.</i> 2011	x		<i>P. composti</i> CCUG 59231 (P182)

Spezies nach LPSN		Referenz <sup>LPSN</sup>	Art des Typstammes nach LPSN		Typstamm im Datensatz (ID)
Nr.	Name		Spezies	Subspezies reklassifiziert	
37	<i>P. congelans</i>	Behrendt <i>et al.</i> 2003	x		<i>P. congelans</i> DSM 14939 (P183)
38	<i>P. corrugata</i>	Roberts & Scarlett 1981	x		<i>P. corrugata</i> DSM 7228 (P185)
39	<i>P. costantinii</i>	Munsch <i>et al.</i> 2002	x		<i>P. costantinii</i> LMG 22119 (P186)
40	<i>P. cremoricolorata</i>	Uchino <i>et al.</i> 2002	x		<i>P. cremoricolorata</i> DSM 17059 (P187)
41	<i>P. cuatrocienegasensis</i>	Escalante <i>et al.</i> 2009	x		<i>P. cuatrocienegasensis</i> CIP 109853 (P189)
42	<i>P. deceptionensis</i>	Carrión <i>et al.</i> 2011	x		<i>P. deceptionensis</i> DSM 26521 (P010)
43	<i>P. delhiensis</i>	Prakash <i>et al.</i> 2007	x		<i>P. delhiensis</i> CCM 7361 (P190)
44	<i>P. donghuensis</i>	Gao <i>et al.</i> 2015	x		<i>P. donghuensis</i> HYS (P191)
45	<i>P. duriflava</i>	Liu <i>et al.</i> 2008	x		<i>P. duriflava</i> KCTC 22129 (P446) <sup>MLSA</sup>
46	<i>P. endophytica</i>	Ramírez-Bahena <i>et al.</i> 2015	x		<i>P. endophytica</i> BSTT44 (P192)
47	<i>P. entomophila</i>	Mulet <i>et al.</i> 2012	x		<i>P. entomophila</i> L48 (P193)
48	<i>P. extremaustralis</i>	López <i>et al.</i> 2010	x		<i>P. extremaustralis</i> DSM 17835 (P011)
49	<i>P. extremorientalis</i>	Ivanova <i>et al.</i> 2002	x		<i>P. extremorientalis</i> LMG 19695 (P194)
50	<i>P. ficuserectae</i>	Goto 1983	x		<i>P. ficuserectae</i> ICMP 7848 (P195)
51	<i>P. flavescens</i>	Hildebrand <i>et al.</i> 1994	x		<i>P. flavescens</i> NBRC 103044 (P196)
52	<i>P. flexibilis</i>	(Hespell 1977), Shin <i>et al.</i> 2016	x		<i>P. flexibilis</i> CGMCC 1.1365 (P197)
53	<i>P. fluorescens</i>	Migula 1895	x		<i>P. fluorescens</i> DSM 50090 (P012)
54	<i>P. formosensis</i>	Lin <i>et al.</i> 2013	x		<i>P. formosensis</i> JCM 18415 (P231)
55	<i>P. fragi</i>	Eichholz 1902	x		<i>P. fragi</i> B25 (P233)
56	<i>P. frederiksbergensis</i>	Andersen <i>et al.</i> 2000	x		<i>P. frederiksbergensis</i> DSM 13022 (P447) <sup>MLSA</sup>
57	<i>P. fulva</i>	Iizuka & Komagata 1963	x		<i>P. fulva</i> NBRC 16637 (P237)
58	<i>P. fuscovaginae</i>	ex Tanii <i>et al.</i> 1976	x		<i>P. fuscovaginae</i> LMG 2158 (P239)
59	<i>P. gelidicola</i>	Kadota 1951	x		---
60	<i>P. gessardii</i>	Verhille <i>et al.</i> 1999	x		<i>P. gessardii</i> DSM 17152 (P026)
61	<i>P. glareae</i>	Romanenko <i>et al.</i> 2015	x		---
62	<i>P. graminis</i>	Behrendt <i>et al.</i> 1999	x		<i>P. graminis</i> DSM 11363 (P241)
63	<i>P. granadensis</i>	Pascual <i>et al.</i> 2015	x		<i>P. granadensis</i> LMG 27940 (P242)
64	<i>P. grimontii</i>	Báida <i>et al.</i> 2002	x		<i>P. grimontii</i> DSM 17515 (P028)
65	<i>P. guangdongensis</i>	Yang <i>et al.</i> 2013	x		<i>P. guangdongensis</i> CCTCC 2012022 (P243)
66	<i>P. guariconensis</i>	Toro <i>et al.</i> 2013	x		<i>P. guariconensis</i> LMG 27394 (P244)
67	<i>P. guguanensis</i>	Liu <i>et al.</i> 2013	x		<i>P. guguanensis</i> JCM 18416 (P245)
68	<i>P. guineae</i>	Bozal <i>et al.</i> 2007	x		<i>P. guineae</i> LMG 24016 (P246)
69	<i>P. helleri</i>	von Neubeck <i>et al.</i> 2016	x		<i>P. helleri</i> WS 4917 (P029)
70	<i>P. helmanticensis</i>	Ramírez <i>et al.</i> 2014	x		<i>P. helmanticensis</i> OHA11 (P448) <sup>MLSA</sup>
71	<i>P. hussainii</i>	Hameed <i>et al.</i> 2014	x		<i>P. hussainii</i> JCM 19513 (P247)
72	<i>P. indica</i>	Pandey <i>et al.</i> 2002	x		<i>P. indica</i> NBRC 103045 (P248)
73	<i>P. japonica</i>	Pungrasmi <i>et al.</i> 2008	x		<i>P. japonica</i> NBRC 103040 (P249)
74	<i>P. jessenii</i>	Verhille <i>et al.</i> 1999	x		<i>P. jessenii</i> CIP 105274 (P449) <sup>MLSA</sup>
75	<i>P. jinjuensis</i>	Kwon <i>et al.</i> 2003	x		<i>P. jinjuensis</i> NBRC 103047 (P250)
76	<i>P. kilonensis</i>	Sikorski <i>et al.</i> 2001	x		<i>P. kilonensis</i> DSM 13647 (P251)
77	<i>P. knackmussii</i>	Stolz <i>et al.</i> 2007	x		<i>P. knackmussii</i> B13 (P252)
78	<i>P. koreensis</i>	Kwon <i>et al.</i> 2003	x		<i>P. koreensis</i> DSM 16610 (P031)
79	<i>P. kunmingensis</i>	Xie <i>et al.</i> 2014	x		<i>P. kunmingensis</i> DSM 25974 (P253)

Spezies nach LPSN		Referenz <sup>LPSN</sup>	Art des Typstammes nach LPSN			Typstamm im Datensatz (ID)
Nr.	Name		Spezies	Subspezies	reklassifiziert	
80	<i>P. kuykendallii</i>	Hunter & Manter 2012	x			<i>P. kuykendallii</i> NRRL B-59562 (P254)
81	<i>P. libanensis</i>	Dabboussi <i>et al.</i> 1999	x			<i>P. libanensis</i> DSM 17149 (P034)
82	<i>P. lini</i>	Delorme <i>et al.</i> 2002	x			<i>P. lini</i> DSM 16768 (P035)
83	<i>P. linyingensis</i>	He <i>et al.</i> 2015	x			<i>P. linyingensis</i> LMG 25967 (P255)
84	<i>P. litoralis</i>	Pascual <i>et al.</i> 2012	x			<i>P. litoralis</i> 2SM5 (P256)
85	<i>P. lundensis</i>	Molin <i>et al.</i> 1986	x			<i>P. lundensis</i> DSM 6252 (P036)
86	<i>P. lurida</i>	Behrendt <i>et al.</i> 2007	x			<i>P. lurida</i> P513-18 (P450) <sup>MLSA</sup>
87	<i>P. lutea</i>	Peix <i>et al.</i> 2004	x			<i>P. lutea</i> DSM 17257 (P257)
88	<i>P. luteola</i>	Kodama <i>et al.</i> 1985	x			<i>P. luteola</i> NBRC 103146 (P258)
89	<i>P. mandelii</i>	Verhille <i>et al.</i> 1999	x			<i>P. mandelii</i> DSM 17967 (P050)
90	<i>P. marginalis</i>	Brown 1918	x			<i>P. marginalis</i> DSM 13124 (P052)
91	<i>P. marincola</i>	Romanenko <i>et al.</i> 2008	x			<i>P. marincola</i> JCM 14761 (P261)
92	<i>P. matsuisoli</i>	Lin <i>et al.</i> 2015	x			---
93	<i>P. mediterranea</i>	Catara <i>et al.</i> 2002	x			<i>P. mediterranea</i> CFBP 5447 (P262)
94	<i>P. meliae</i>	Ogimi 1981	x			<i>P. meliae</i> CFBP 3225 (P263)
95	<i>P. mendocina</i>	Palleroni 1970	x			<i>P. mendocina</i> NBRC 14162 (P266)
96	<i>P. meridiana</i>	Reddy <i>et al.</i> 2004	x			<i>P. meridiana</i> CIP 108465 (P451) <sup>MLSA</sup>
97	<i>P. migulae</i>	Verhille <i>et al.</i> 1999	x			<i>P. migulae</i> NBRC 103157 (P271)
98	<i>P. mohnii</i>	Cámara <i>et al.</i> 2007	x			<i>P. mohnii</i> DSM 18327 (P272)
99	<i>P. montellii</i>	Elomari <i>et al.</i> 1997	x			<i>P. montellii</i> NBRC 103158 (P273)
100	<i>P. moorei</i>	Cámara <i>et al.</i> 2007	x			<i>P. moorei</i> CCUG 53114 (P452) <sup>MLSA</sup>
101	<i>P. moraviensis</i>	Tvrzová <i>et al.</i> 2006	x			<i>P. moraviensis</i> DSM 16007 (P453) <sup>MLSA</sup>
102	<i>P. mosselii</i>	Dabboussi <i>et al.</i> 2002	x			<i>P. mosselii</i> DSM 17497 (P277)
103	<i>P. mucidolens</i>	Levine & Anderson 1932	x			<i>P. mucidolens</i> NBRC 103159 (P279)
104	<i>P. nitritireducens</i>	Wang <i>et al.</i> 2015	x			---
105	<i>P. nitroreducens</i>	Iizuka & Komagata 1964	x			<i>P. nitroreducens</i> DSM 14399 (P053)
106	<i>P. oceani</i>	Wang & Sun 2016	x			---
107	<i>P. oleovorans</i> (ssp. <i>lubricantis</i> )	Saha <i>et al.</i> 2010		x		<i>P. oleovorans</i> (ssp. <i>lubricantis</i> ) RS1 (P282)
107	<i>P. oleovorans</i> (ssp. <i>oleovorans</i> )	Lee & Chandler 1941	x	x		<i>P. oleovorans</i> (ssp. <i>oleovorans</i> ) NBRC 13583 (P283)
107	<i>P. oleovorans</i>	Stanier 1966			x	<i>P. pseudoalcaligenes</i> NBRC 14167 (P308)
108	<i>P. orientalis</i>	Dabboussi <i>et al.</i> 2002	x			<i>P. orientalis</i> DSM 17489 (P055)
109	<i>P. oryzihabitans</i>	Kodama <i>et al.</i> 1985	x			<i>P. oryzihabitans</i> DSM 6835 (P056)
110	<i>P. otitidis</i>	Clark <i>et al.</i> 2006	x			<i>P. otitidis</i> DSM 17224 (P287)
111	<i>P. pachastrellae</i>	Romanenko <i>et al.</i> 2005	x			<i>P. pachastrellae</i> CCUG 46540 (P289)
112	<i>P. palleroniana</i>	Gardan <i>et al.</i> 2002	x			<i>P. palleroniana</i> LMG 23076 (P454) <sup>MLSA</sup>
113	<i>P. panacis</i>	Park <i>et al.</i> 2005	x			<i>P. panacis</i> DSM 18529 (P058)
114	<i>P. panipatensis</i>	Gupta <i>et al.</i> 2008	x			<i>P. panipatensis</i> CCM 7469 (P290)
115	<i>P. parafulva</i>	Uchino <i>et al.</i> 2002	x			<i>P. parafulva</i> DSM 17004 (P292)
116	<i>P. pelagia</i>	Hwang <i>et al.</i> 2009	x			<i>P. pelagia</i> CL-AP6 (P293)
117	<i>P. peli</i>	Vanparys <i>et al.</i> 2006	x			<i>P. peli</i> DSM 17833 (P061)
118	<i>P. pertucinogena</i>	Kawai & Yabuuchi 1975	x			<i>P. pertucinogena</i> LMG 1874 (P455) <sup>MLSA</sup>
119	<i>P. plecoglossicida</i>	Nishimori <i>et al.</i> 2000	x			<i>P. plecoglossicida</i> NBRC 103162 (P296)
120	<i>P. poae</i>	Behrendt <i>et al.</i> 2003	x			<i>P. poae</i> DSM 14936 (P063)

Spezies nach LPSN		Referenz <sup>LPSN</sup>	Art des Typstammes nach LPSN		Typstamm im Datensatz (ID)
Nr.	Name		Spezies	Subspezies reklassifiziert	
121	<i>P. pohangensis</i>	Weon <i>et al.</i> 2006	x		<i>P. pohangensis</i> DSM 17875 (P300)
122	<i>P. populi</i>	Anwar <i>et al.</i> 2016	x		---
123	<i>P. prosekii</i>	Kosina <i>et al.</i> 2014	x		<i>P. prosekii</i> LMG 26867 (P301)
124	<i>P. protegens</i>	Ramette <i>et al.</i> 2012	x		<i>P. protegens</i> CHAO (P303)
125	<i>P. proteolytica</i>	Reddy <i>et al.</i> 2004	x		<i>P. proteolytica</i> DSM 15321 (P066)
126	<i>P. psychrophila</i>	Yumoto <i>et al.</i> 2002	x		<i>P. psychrophila</i> DSM 17535 (P083)
127	<i>P. psychrotolerans</i>	Hauser <i>et al.</i> 2004	x		<i>P. psychrotolerans</i> DSM 15758 (P084)
128	<i>P. punonensis</i>	Ramos <i>et al.</i> 2013	x		<i>P. punonensis</i> CECT 8089 (P311)
129	<i>P. putida</i>	Trevisan 1889	x		<i>P. putida</i> NBRC 14164 (P324)
130	<i>P. reinekei</i>	Cámara <i>et al.</i> 2007	x		<i>P. reinekei</i> MT-1 (P340)
131	<i>P. resinovorans</i>	Delaporte <i>et al.</i> 1961	x		<i>P. resinovorans</i> DSM 21078 (P341)
132	<i>P. rhizosphaerae</i>	Peix <i>et al.</i> 2003	x		<i>P. rhizosphaerae</i> DSM 16299 (P343)
133	<i>P. rhodesiae</i>	Coroler <i>et al.</i> 1997,	x		<i>P. rhodesiae</i> DSM 14020 (P085)
134	<i>P. sabulinigri</i>	Kim <i>et al.</i> 2009	x		<i>P. sabulinigri</i> JCM 14963 (P345)
135	<i>P. sagittaria</i>	Lin <i>et al.</i> 2013	x		<i>P. sagittaria</i> JCM 18195 (P346)
136	<i>P. salegens</i>	Amoozegar <i>et al.</i> 2014	x		<i>P. salegens</i> CECT 8338 (P347)
137	<i>P. salina</i>	Zhong <i>et al.</i> 2015	x		<i>P. salina</i> XCD-X85 (P456) <sup>MLSA</sup>
138	<i>P. salomonii</i>	Gardan <i>et al.</i> 2002	x		<i>P. salomonii</i> ICMP 14252 (P348)
139	<i>P. saponiphila</i>	Lang <i>et al.</i> 2012	x		<i>P. saponiphila</i> DSM 9751 (P349)
140	<i>P. savastanoi</i>	Janse 1982	x		<i>P. savastanoi</i> ICMP 4352 (P350)
141	<i>P. segetis</i>	Park <i>et al.</i> 2006	x		<i>P. segetis</i> IMSNU 14100 (P457) <sup>MLSA</sup>
142	<i>P. seleniipraecipitans</i>	Hunter & Manter 2011	x		<i>P. seleniipraecipitans</i> LMG 25475 (P351)
143	<i>P. simiae</i>	Vela <i>et al.</i> 2006	x		<i>P. simiae</i> CCUG 50988 (P353)
144	<i>P. soli</i>	Pascual <i>et al.</i> 2015	x		<i>P. soli</i> LMG 27941 (P357)
145	<i>P. straminea</i>	Iizuka & Komagata 1963	x		<i>P. straminea</i> JCM 2783 (P364)
146		Wolterink <i>et al.</i> 2002		x	<i>P. chloritidismutans</i> AW-1 (P166)
146	<i>P. stutzeri</i>	Lehmann & Neumann 1896	x		<i>P. stutzeri</i> ATCC 17588 (P368)
146		ex ZoBell & Upham 1944		x	<i>P. perfectomarina</i> ATCC 14405 (P294)
147	<i>P. synxantha</i>	Ehrenberg 1840	x		<i>P. synxantha</i> DSM 18928 (P091)
148	<i>P. syringae</i> (ssp. <i>syringae</i> )	van Hall 1902	x	x	<i>P. syringae</i> DSM 10604 (P380)
149	<i>P. taeaanensis</i>	Lee <i>et al.</i> 2010	x		<i>P. taeaanensis</i> MS-3 (P423)
150	<i>P. taetrolens</i>	Haynes 1957	x		<i>P. taetrolens</i> DSM 21104 (P092)
151	<i>P. taiwanensis</i>	Wang <i>et al.</i> 2010	x		<i>P. taiwanensis</i> DSM 21245 (P424)
152	<i>P. thermotolerans</i>	Manaiia & Moore 2002	x		<i>P. thermotolerans</i> DSM 14292 (P425)
153	<i>P. thivervalensis</i>	Achouak <i>et al.</i> 2000	x		<i>P. thivervalensis</i> DSM 13194 (P427)
154	<i>P. tolaasii</i>	Paine 1919	x		<i>P. tolaasii</i> NCPPB 2192 (P428)
155	<i>P. toyotomiensis</i>	Hirota <i>et al.</i> 2011	x		<i>P. toyotomiensis</i> JCM 15604 (P429)
156	<i>P. tremae</i>	Gardan <i>et al.</i> 1999	x		<i>P. tremae</i> ICMP 9151 (P430)
157	<i>P. trivialis</i>	Behrendt <i>et al.</i> 2003	x		<i>P. trivialis</i> DSM 14937 (P093)
158	<i>P. tuomuerensis</i>	Xin <i>et al.</i> 2009	x		<i>P. tuomuerensis</i> JCM 14085 (P431)
159	<i>P. turukhanskensis</i>	Korshunova <i>et al.</i> 2016	x		<i>P. turukhanskensis</i> IB1.1 (P458) <sup>MLSA</sup>
160	<i>P. umsongensis</i>	Kwon <i>et al.</i> 2003	x		<i>P. umsongensis</i> LMG 21317 (P459) <sup>MLSA</sup>
161	<i>P. vancouverensis</i>	Mohn <i>et al.</i> 1999	x		<i>P. vancouverensis</i> ATCC 700688 (P460) <sup>MLSA</sup>

Spezies nach LPSN		Referenz <sup>LPSN</sup>	Art des Typstammes nach LPSN		Typstamm im Datensatz (ID)
Nr.	Name		Spezies	Subspezies reklassifiziert	
162	<i>P. veronii</i>	Elomari <i>et al.</i> 1996	x		<i>P. veronii</i> DSM 11331 (P094)
163	<i>P. viridiflava</i>	Burkholder 1930	x		<i>P. viridiflava</i> DSM 6694 (P434)
164	<i>P. vranovensis</i>	Tvrzová <i>et al.</i> 2006	x		<i>P. vranovensis</i> DSM 16006 (P437)
165	<i>P. weihenstephanensis</i>	von Neubeck <i>et al.</i> 2016	x		<i>P. weihenstephanensis</i> WS 4993 (P099)
166	<i>P. xanthomarina</i>	Romanenko <i>et al.</i> 2005	x		<i>P. xanthomarina</i> DSM 18231 (P438)
167	<i>P. xiamenensis</i>	Lai & Shao 2008	x		---
168	<i>P. xinjiangensis</i>	Liu <i>et al.</i> 2009	x		---
169	<i>P. yamanorum</i>	Arnau <i>et al.</i> 2015	x		<i>P. yamanorum</i> LMG 27247 (P440)
170	<i>P. zeshuii</i>	Feng <i>et al.</i> 2012	x		<i>P. zeshuii</i> KACC 15471 (P441)
171	<i>P. zhaodongensis</i>	Zhang <i>et al.</i> 2015	x		---
	<i>P. "canadensis"</i>	Tambong <i>et al.</i> 2017			<i>P. "canadensis"</i> 2-92 (P161)
	<i>P. "cerasi"</i>	Kalužna <i>et al.</i> 2016			<i>P. "cerasi"</i> 58 (P165)
	<i>P. "lactis"</i>	von Neubeck <i>et al.</i> 2017			<i>P. "lactis"</i> WS 4992 (P033)
	<i>P. "oryzae"</i>	Yu <i>et al.</i> 2013			<i>P. "oryzae"</i> KCTC 32247 (P286)
	<i>P. "paralactis"</i>	von Neubeck <i>et al.</i> 2017			<i>P. "paralactis"</i> WS 4672 (P060)
	<i>P. "sihuiensis"</i>	Wu <i>et al.</i> 2014			<i>P. "sihuiensis"</i> KCTC 32246 (P352)
	<i>P. "versuta"</i>	See-Too <i>et al.</i> 2017			<i>P. "versuta"</i> L10.10 (P433)

<sup>LPSN</sup>, Referenz wurde von der LPSN [67] (<http://www.bacterio.net/>) übernommen

<sup>MLSA</sup>, Lediglich MLSA Sequenzdaten zum entsprechenden Stamm im Datensatz vorhanden (siehe P442-P460 in Tabelle 30)

## XII. PUBLIKATIONEN UND KONFERENZBEITRÄGE

### 1. Publikationen

Teile der vorliegenden Arbeit wurden bereits vorab publiziert in

Huptas C., Scherer S., Wenning M.: **Optimized Illumina PCR-free library preparation for bacterial whole genome sequencing and analysis of factors influencing *de novo* assembly.** *BMC Res Notes* 2016, 9:269. DOI: 10.1186/s13104-016-2072-9

Weitere Publikationen mit persönlicher Beteiligung sind

Böhm M.E., Huptas C., Krey V.M., Scherer S.: **Massive horizontal gene transfer, strictly vertical inheritance and ancient duplications differentially shape the evolution of *Bacillus cereus* enterotoxin operons *hbl*, *cytK* and *nhe*.** *BMC Evol Biol* 2015, 15:246. DOI: 10.1186/s12862-015-0529-4

Böhm M.E., Huptas C., Krey V.M., Scherer S.: **Draft Genome Sequence of *Bacillus cytotoxicus* CVUAS 2833, a Very Close Relative to Type Strain NVH 391-98 Isolated from a Different Location.** *Genome Announc* 2015, 3(4). DOI: 10.1128/genomeA.00901-15

Müller A., Huptas C., Wenning M., Schmidt H., Weiss A.: **Draft Genome Sequence of *Staphylococcus carnosus* subsp. *utilis* LTH 7013, Isolated from South Tyrolean Ham.** *Genome Announc* 2015, 3(3). DOI: 10.1128/genomeA.00456-15

Fellner L., Huptas C., Simon S., Mühlig A., Scherer S., Neuhaus K.: **Draft Genome Sequences of Three European Laboratory Derivatives from Enterohemorrhagic *Escherichia coli* O157:H7 Strain EDL933, Including Two Plasmids.** *Genome Announc* 2016, 4(2). DOI: 10.1128/genomeA.01331-15

Loscar M.E., Huptas C., Wenning M., Sieber V., Schmid J.: **Draft Genome Sequence of *Lysinibacillus xylanilyticus* SR-86.** *Genome Announc* 2016, 4(6). DOI: 10.1128/genomeA.01317-16

Schmid J., Huptas C., Wenning M.: **Draft Genome Sequence of the Xanthan Producer *Xanthomonas campestris* LMG 8031.** *Genome Announc* 2016, 4(5). DOI: 10.1128/genomeA.01069-16

von Neubeck M., Huptas C., Glück C., Krewinkel M., Stoeckel M., Stressler T., Fischer L., Hinrichs J., Scherer S., Wenning M.: ***Pseudomonas helleri* sp. nov. and *Pseudomonas weihenstephanensis* sp. nov., isolated from raw cow's milk.** *Int J Syst Evol Microbiol* 2016, 66(3):1163-1173. DOI: 10.1099/ijsem.0.000852

Ferrari E., Walter M.C., Huptas C., Scherer S., Müller-Herbst S.: **Complete Circular Genome Sequence and Temperature Independent Adaptation to Anaerobiosis of *Listeria weihenstephanensis* DSM 24698.** *Front Microbiol* 2017, 8:1672. DOI: 10.3389/fmicb.2017.01672

von Neubeck M., Huptas C., Glück C., Krewinkel M., Stoeckel M., Stressler T., Fischer L., Hinrichs J., Scherer S., Wenning M.: ***Pseudomonas lactis* sp. nov. and *Pseudomonas paralactis* sp. nov., isolated from bovine raw milk.** *Int J Syst Evol Microbiol* 2017, 67(6):1656-1664. DOI: 10.1099/ijsem.0.001836

Wenning M., Breitenwieser F., Huptas C., von Neubeck M., Busse H.-J., Scherer S.: ***Lactibacter flavus* gen. nov., sp. nov., a novel bacterium of the family Propionibacteriaceae isolated from raw milk and dairy products.** (In Revision, Manuskript Referenz-Nr.: IJSEM-D-18-00202)

## 2. Konferenzbeiträge

Huptas C., Scherer S., Wenning M.: **Use of whole-genome sequences for delineation of *Pseudomonas* species.** *EMBO Conference on Microbiology after the genomics revolution: Genomes 2014*, Paris, Frankreich, 24.-27.06.2014 (Poster)

Fellner L., Landstorfer R., Huptas C., Simon S., Marx H., Mühlig A., Küster B., Scherer S., Neuhaus K.: **Seventy novel small proteins of *Escherichia coli* O157:H7 EDL933 and a complete revision of its genetic makeup.** *EMBO Conference on Microbiology after the genomics revolution: Genomes 2014*, Paris, Frankreich, 24.-27.06.2014 (Poster)

Huptas C., Scherer S., Wenning M.: **Optimized protocols for Illumina PCR-free whole genome sequencing and evaluation of assembly performance.** 4. *Gemeinsame Konferenz von DGHM und VAAM*, Dresden, Deutschland, 05.-08.10.2014 (Poster)

Böhm M.E., Huptas C., Krey V.M., Scherer S.: **Genomic insights into species affiliation and propagation of virulence determinants in *B. cereus* sensu lato.** *VAAM-Jahrestagung 2015*, Marburg/Lahn, Deutschland, 01.-04.03.2015 (Poster)

Lücking G., von Neubeck M., Huptas C., Scherer S., Wenning M.: **Proteolytic activity of *Pseudomonas* spp. isolated from raw milk.** *5. Gemeinsame Tagung von DGHM und VAAM*, Würzburg, Deutschland, 05.-08.3.2017 (Poster)

Huptas C., Kreitmeier M., Wenning M., Scherer S.: **Delineation of *Pseudomonas* species based on genomic sequences.** *16th International Conference on Pseudomonas*, Liverpool, Vereinigtes Königreich, 05.-09.09.2017 (Poster)

Maier C., Lücking G., von Neubeck M., Huptas C., Scherer S., Wenning M.: **AprA peptidase production of *Pseudomonas* spp. isolated from raw milk.** *Jahrestagung 2018 der Vereinigung für Allgemeine und Angewandte Mikrobiologie*, Wolfsburg, Deutschland, 15.04.-18.04.2018 (Poster)