



Implementierung eines 1-step-full-length-16S rRNA-Sequenzierungsprotokolls bei der NGSP

Tatjana, Maitland BMA 20-23

Biomedizinische Analytik HF

1. Zusammenfassung

Bei der NGSP sollte ein 1-step full-length-16S rRNA-Protokoll von Pacbio implementiert werden. Der 16S-Locus wirkt wie ein Barcode zur Unterscheidung mikrobieller Taxa und kann für die Klassifikation von Bakterien verwendet werden. Das Protokoll wurde mit einem Batch aus 119 DNA-Proben durchgeführt. Dafür wurde die Long read SMRT bell Technologie von Pacbio verwendet. Dies ist eine third generation sequencing methode die es erlaubt sehr lange DNA-Fragmente zu sequenzieren. Nach der Sequenzierung des Projekts wurde der Erfolg der Implementierung in Bezug auf sequencing Output (Ziel über 10'000 Polymerasereads pro Probe) und der Analyse der verwendeten Kontrollen beurteilt. Es wurden 2 negative Kontrollen genutzt um Kontaminationen aus Reagenzien oder fremder DNA zu detektieren. Und 2 positive Kontrollen (Com.control und Log Control) um bias, Fehler und detektionslimits im workflow zu erkennen. Im Allgemeinen funktionierte das Experiment und die Mehrheit der Proben wurde erfolgreich mit mehr als 10'000 Reads des 1.5KB 16S rRNA-Gens sequenziert. Die Kontrollen wurden von der Bioinformatikerin der NGSP genauer analysiert. Es konnten alle Arten identifiziert werden, die im Rahmen der Standard Com.control angegeben sind. Ebenfalls wurden die ersten 6 Bakterienarten von den 8 in der com. Log. Kontrolle vorhandenen Arten identifiziert. Die letzten 2 Bakterienarten, die in vernachlässigbaren Mengen vorhanden waren, konnten wir nicht nachweisen. Tatsächlich kann die Zymo-Forschung selbst die letzten 3 Bakterienarten mit derselben Bioinformatik-Pipeline (Dada2) nicht identifizieren. Bei einer Negativkontrolle, der NTC, wurde 1 Bakterienspezies (namens *C. acnes*) identifiziert. Jedoch gibt es nur sehr wenige reads und diese Identifizierung ist mit geringer Konfidenz verbunden deshalb war diese Kontrolle gültig. Die Implementierung des neuen Protokoll war erfolgreich.

2. Einleitung

Die Next generation sequencing Plattform (NGSP) ist eine Forschungsserviceeinrichtung der Universität Bern, die Forschende der Uni Bern, des Inselspitals und weitere Forschungsgruppen in der ganzen Schweiz bei NGS-Projekten unterstützt. NGS hat die Life-Sciences-Forschung revolutioniert, und ein explodierendes Forschungsfeld ist die Mikrobiomik. Mikrobiomik ist die Wissenschaft der kollektiven Charakterisierung und Quantifizierung von Molekülen, die für die Struktur, Funktion und Dynamik einer mikrobiellen Gemeinschaft verantwortlich sind [1]. Traditionell erfolgt die Identifizierung von Bakterienarten aus komplexen Proben durch Kultivierung, MALDI-TOF und biochemische Tests, die jedoch ungenau und langsam sein können [2]. Die beiden am häufigsten verwendeten NGS-basierten Ansätze zur mikrobiellen Identifizierung basieren auf der Genamplifikation des 16S rRNA-Gens oder der Shotgun-Metagenomik [3]. 16S rRNA steht für 16S ribosomale Ribonukleinsäure (rRNA) und ist ein wichtiger Bestandteil der kleinen Untereinheit (SSU) der prokaryotischen Ribosomen. Das 16S rRNA-Gen besteht sowohl aus hochkonservierten als auch aus hypervariablen (V1-V9) Regionen [4].

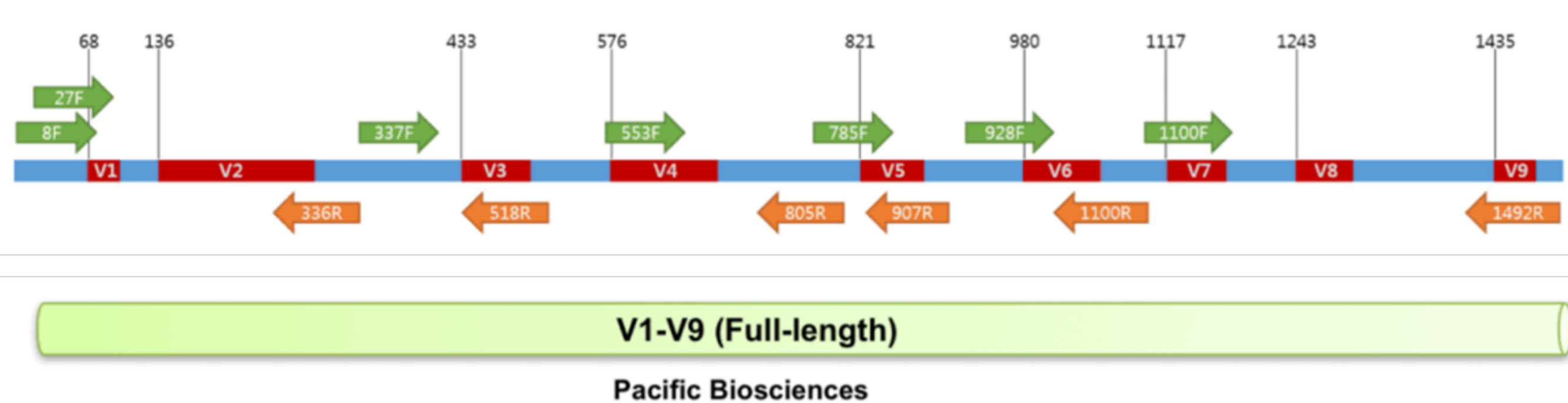


Abb. 1: 16S rRNA-Gen-Diagramm. Blau steht für die universellen Sequenzen und die roten hypervariablen Regionen. Die grünen Pfeile sind häufig verwendete Primer für die Amplifikation des Gens. Bei dieser einstufigen PCR 16S-Methode wird das gesamte 16S rRNA-Gen einschließlich der Regionen V1-V9 amplifiziert.

Die konservierten Regionen dienen als Primer-Bindungsstellen für die PCR-Amplifikation der variablen Regionen, die Sequenzen enthalten, die zur bakteriellen Identifizierung und Klassifizierung verwendet werden können (Abb. 1). Wenn alle 9 variablen Regionen mit Hilfe von Long-Read-Technologien sequenziert werden können, wird die phylogenetische Auflösung erhöht und die Bakterienarten können genauer auf Artebene identifiziert werden [5]. Bis vor kurzem wurde bei 16S rRNA-Seq ein 2-step PCR-Ansatz verwendet. Die erste PCR verwendet einen universellen 16S-Primer mit Primerende und die zweite PCR wird an den universellen Enden angebaut und fügt einen eindeutigen Barcode hinzu. Danach werden die Barcode-16S-Genamplikone gepoolt, eine NGS-Library aufgebaut und die Hochdurchsatz-Sequenzierung durchgeführt. Es wäre schneller und würde die Gefahr einer Kontamination verringern, wenn es nur 1 PCR-Schritt gäbe.

3. Ziele und Fragestellungen

Ziel 1: Implementierung eines 1-step-full-length 16s rRNA Sequenzierungsprotokolls bei der NGSP auf dem Sequel Ile von Pacbio.

Fragestellung zu Ziel 1: War die Implementierung des Protokolls erfolgreich in Bezug des sequencing output (Ziel über 8'000 Polymerasereads pro Probe) und der Analyse von Negativ- und Positivkontrollen?

4. Material, Methodik, Vorgehen

Die exakte Methodik ist in der Abb. 2 dargestellt.

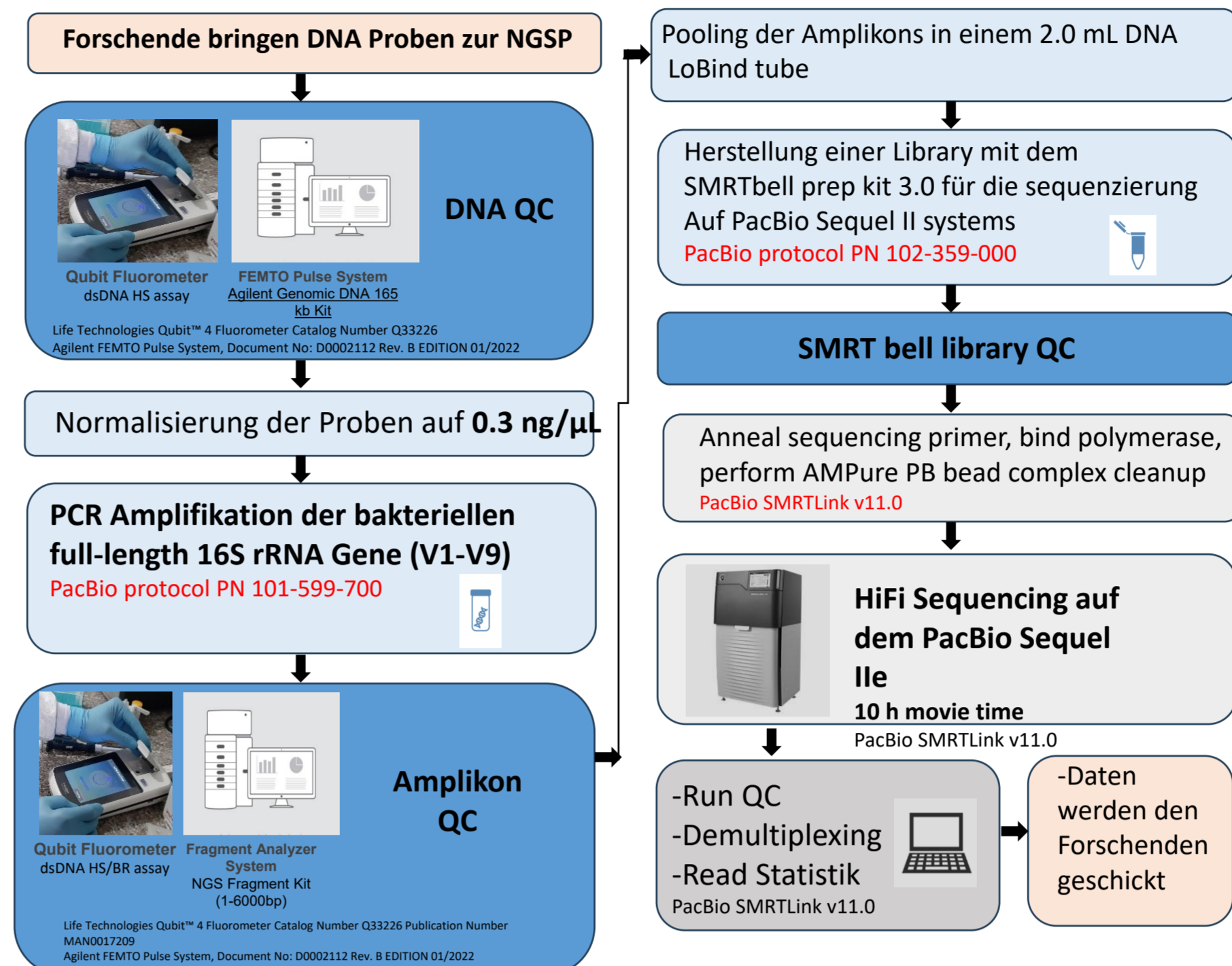


Abb. 2, links: Full-length 16S rRNA Gen Sequenzierung, service method übersicht vom Probeneingang bis zur Datenfreigabe.

Species	Genomid. DNA	Species	Genomid. DNA
<i>Pseudomonas aeruginosa</i>	12	<i>Listeria monocytogenes</i>	89.1
<i>Escherichia coli</i>	12	<i>Pseudomonas aeruginosa</i>	8.9
<i>Salmonella enterica</i>	12	<i>Bacillus subtilis</i>	0.99
<i>Lactobacillus fermentum</i>	12	<i>Escherichia coli</i>	0.089
<i>Enterococcus faecalis</i>	12	<i>Salmonella enterica</i>	0.089
<i>Staphylococcus aureus</i>	12	<i>Lactobacillus fermentum</i>	0.0089
<i>Listeria monocytogenes</i>	12	<i>Enterococcus faecalis</i>	0.00089
<i>Bacillus subtilis</i>	12	<i>Staphylococcus aureus</i>	0.000089

Abb. 3, Oben: A. ZymoBIOMICS Microbial Community DNA Standard Kontrolle (D6305, Zymo Research) mit 8 definierten Bakterienarten in gleicher Menge. B. ZymoBIOMICS Microbial Community Standard II (Log Distribution) Kontrolle (D6311, Zymo Research) mit 8 definierten Bakterienarten gemischt, um eine logarithmische Verteilung der Abundanz zu erhalten.

Die PCR wurde mit 2 Negativ-Kontrollen (NTC und Elutionbuffer) durchgeführt um Kontaminationen aus Reagenzien oder fremder DNA zu detektieren. Für die PCR wurden auch 2 Positivkontrollen eingesetzt. Diese wurden eingesetzt um bias, Fehler und detektionslimits im workflow zu erkennen. Die Komposition dieser Kontrollen sind in der Abb.3 sichtbar.

5. Ergebnisse/ Resultate

Allgemein funktionierte das Experiment und die Mehrheit der Proben wurde erfolgreich mit mehr als 10'000 Reads des 1.5KB 16S rRNA-Gens sequenziert. (Siehe Abb.4) Die Bioinformatikerin der NGSP hat alle Kontrollen genauer analysiert. Wir waren in der Lage, alle Arten zu identifizieren, die im Rahmen der Standard Com. control angegeben sind, (nicht in der genauen Häufigkeit, die vom Unternehmen angegeben wurde, aber alle Bakterienarten wurden identifiziert. Es konnten die ersten 5 Bakterienarten von den 8 in der com. Log. Kontrolle vorhandenen Arten identifizieren. Die letzten 3 Bakterienarten, die in vernachlässigbaren Mengen vorhanden waren, konnten nicht nachweisen werden. Tatsächlich kann der Hersteller der Kontrolle selbst die letzten 3 Bakterienarten mit derselben Bioinformatik-Pipeline nicht identifizieren. Bei einer Negativkontrolle, der NTC, wurde 1 Bakterienspezies namens *C. acnes* identifiziert. Jedoch gibt es nur sehr wenige reads und diese Identifizierung ist mit geringer Konfidenz verbunden deshalb war diese Kontrolle gültig.

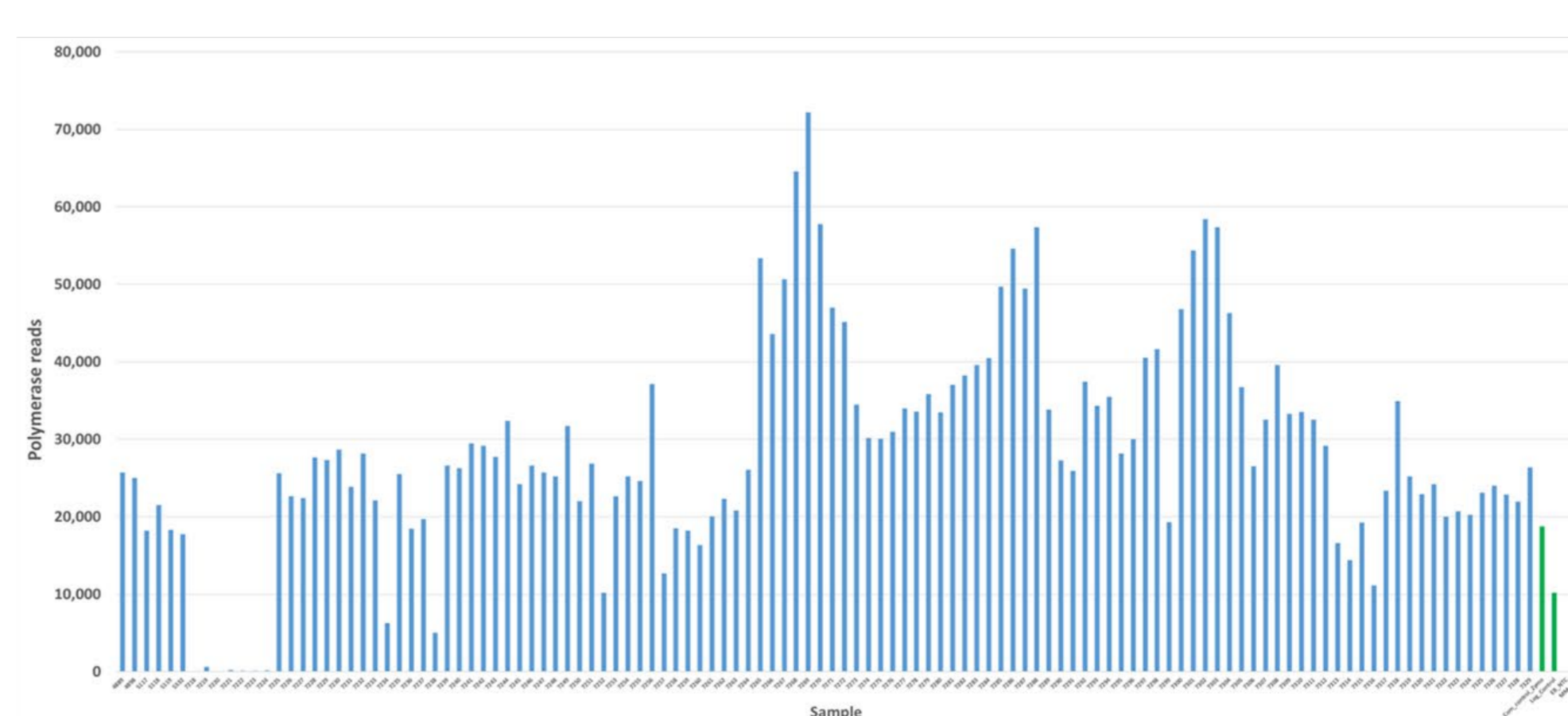


Abb. 4: Anzahl Polymerasereads pro Probe (Balancing) Nach der Sequenzierung

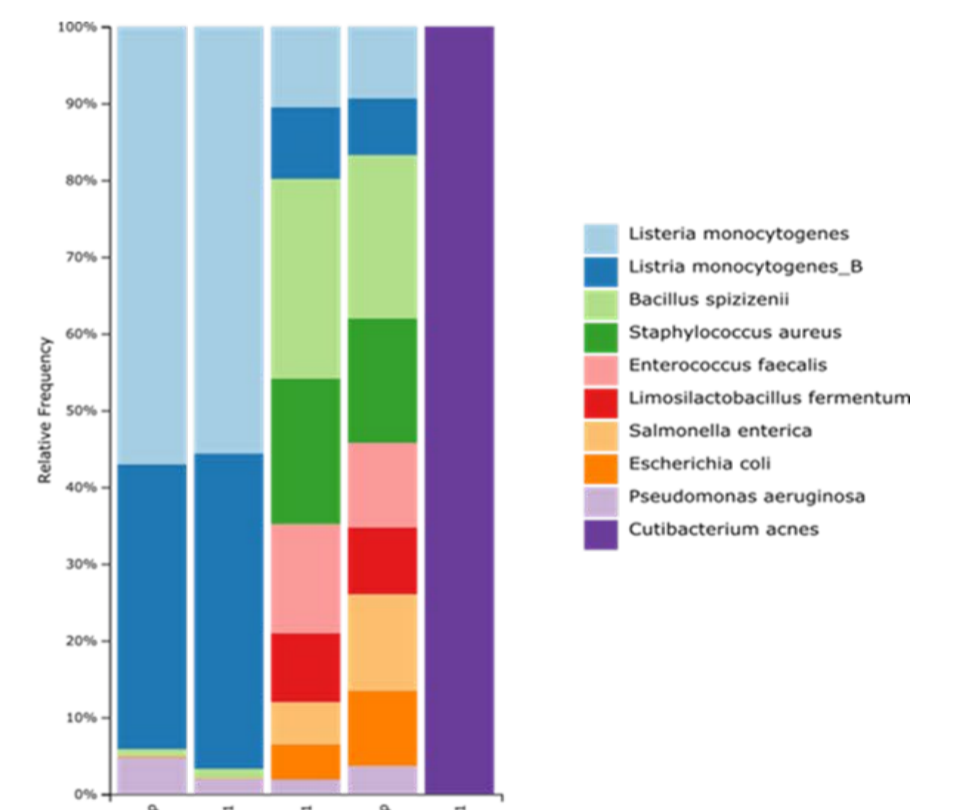


Abb. 5: Bioinformatik-Analyse der im Arbeitsablauf verwendeten Kontrollen. Das Balkendiagramm zeigt die Häufigkeit der Bakterienarten für jede Kontrollprobe an.

6. Diskussion

Im Allgemeinen funktionierte das Experiment und die Mehrheit der Proben wurde erfolgreich mit mehr als 10'000 Reads des 1.5KB 16S rRNA-Gens sequenziert. 9 der 119 Proben generiert $\leq 10'000$ Reads/Prob. Die Proben, die fehlschlagen, wiesen auch die niedrigsten DNA-Input Konzentrationen und die niedrigsten DNA-Qualität auf. Die Verteilung der Polymerasereads war teils sehr unterschiedlich. Trotz einer normalisierten DNA-Eingangsmenge können die Proben unterschiedliche Grade von Komplexitäten oder Verunreinigungen der Bakteriengemeinschaft aufweisen, was sich auf die PCR und die Amplifikation des 16S rRNA-Gens auswirken kann. Daher gab es eine große Bandbreite an PCR-Barcode-Amplikonkonzentrationen. Die Eingesetzten Kontrollen waren valide Angesichts der Ergebnisse bedeutet dies, dass wir während des Workflows keine Fehler oder Verzerrungen eingeführt haben.

Referenzen

[1] Marchesi & Ravel, 2015; [2] Handelsman, 2004; [3] Allaband *et al.*, 2019; [4] Woese 1987; [5] Johnson *et al.*, 2019.

Abbildungen

Abb. 1 16S ribosomale RNA gene sequencing. Diagramm von EzBioCloud Resources for Microbial Bioinformatics
Abb. 2 Übersicht Methodik (eigen Abbildung)

Abb. 3 Diagramm aus den Handbüchern von Zymo Research: Zymo Research ZymoBIOMICS Microbial Community DNA Standard Kontrolle (D6305). B. ZymoBIOMICS Microbial Community Standard II (Log Distribution) Kontrolle (D6311)

Abb.4 Anzahl Polymerasereads pro Probe (Balancing) (Eigene Abbildung)

Abb. 5 Informationen der verwendeten Kontrollen (Eigene Abbildung)