

Aus dem Department für Nutztiere und öffentliches Gesundheitswesen  
der Veterinärmedizinischen Universität Wien

Institut für Lebensmittelsicherheit, Lebensmitteltechnologie und  
öffentliches Gesundheitswesen in der Veterinärmedizin  
Abteilung für Lebensmittelmikrobiologie  
(Leiter: Univ.-Prof. Dr. med. vet. Dipl. ECVPH Martin Wagner)

## **Mikrobielle Veränderungen im Kot von Katzen die an Lymphomen im Verdauungstrakt erkrankt sind**

**Diplomarbeit**

Veterinärmedizinische Universität Wien

vorgelegt von  
Luisa Sontheimer

Wien, im Januar 2022

## Betreuung und Begutachtung

**Betreuerin:** Ass.-Prof. Mag. med. vet. Evelyne Selberherr,  
PhD aus dem Institut für Lebensmittelsicherheit,  
Lebensmitteltechnologie und öffentliches  
Gesundheitswesen in der Veterinärmedizin,  
Abteilung für Lebensmittelmikrobiologie,  
Department für Nutztiere und öffentliches  
Gesundheitswesen in der Veterinärmedizin

**Begutachter/in:** Univ.-Prof. Dr. sc. agr. Barbara Metzler-Zebeli,  
Institut für Physiologie, Pathophysiologie und  
Biophysik,  
Abteilung für Ernährungsphysiologie

## Danksagung

An dieser Stelle möchte ich mich bei all denjenigen Bedanken, die mir während der Anfertigung meiner Diplomarbeit unterstützend und motivierend zur Seite gestanden sind.

Zuallererst möchte ich mich bei meiner Betreuerin Ass.-Prof. Mag. med. vet. Evelyne Selberherr, PHD bedanken. Vielen Dank für Deine großartige Unterstützung, trotz der erschwerten Umstände in Pandemiezeiten. Danke, dass Du immer für mich erreichbar warst und Dir Zeit für alle meine Fragen genommen hast. Dank Deiner tatkräftigen Unterstützung habe ich das Ziel nie aus den Augen verloren.

Des weiteren möchte ich mich bei Mag. med. vet. Stefanie Urimare Wetzels, PhD bedanken. Danke für Deine hervorragende Einführung und Einarbeitung in QIIME2 und die Beantwortung aller Unklarheiten zu diesem bioinformatischen Analysetool.

An dieser Stelle möchte ich mich auch bei Univ.-Prof. Dr.med.vet. Dipl.ECVPH Martin Wagner bedanken. Vielen Dank für die Möglichkeit diese Arbeit bei Ihnen am Institut zu schreiben.

Vielen lieben Dank an die Vetbiobank (Vetmeduni Vienna), Dr. Birgitt Wolfesberger und Prof. Iwan Burgener (Universitätsklinik für Kleintiere, Klinische Abteilung für Interne Medizin Kleintiere), ohne dessen Unterstützung diese Arbeit nicht möglich gewesen wäre.

Ein besonderer Dank geht an meine Eltern, Ute und Dietrich Sontheimer, die mir mein Studium durch Ihre Unterstützung erst ermöglicht haben. Danke, dass Ihr immer an mich glaubt.

# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>1</b>
1.1. Die mikrobielle Vielfalt	1
1.2. Detektionsmethoden von Mikroben	2
1.2.1. Klassische Kultivierungsverfahren	2
1.2.2. Auf PCR-basierende kulturunabhängige Methoden (konventionelle PCR und qPCR)	3
1.2.3. Sequenzierungsmethoden	4
1.3. Bioinformatische Analyse von Hochdurchsatzsequenzierungen	8
1.3.1. Vergleich von bioinformatischen Plattformen	9
1.4. Darmmikrobiota	11
1.5. Hypothese und Ziel dieser Arbeit	13
<b>2. Material und Methoden</b>	<b>14</b>
2.1. Probengewinnung	14
2.2. DNA-Extraktion	14
2.3. DNA-Vermessung	14
2.4. Sequenzierung	15
2.5. Bioinformatische Analyse	15
<b>3. Resultate</b>	<b>27</b>
3.1. Vermessung der DNA	27
3.2. Sequenzstatistik	28
3.3. Die Alpha-Diversität	28
3.4. Auswertung der Taxa-Bar-Plots	29
3.5. Die Beta-Diversität	31
3.6. Die Signifikanzberechnung	31
3.6.1. Alpha- und Beta-Diversität	31
3.7. Die relative Häufigkeit	33
3.8. Die häufigsten ASVs	34
<b>4. Diskussion</b>	<b>36</b>
4.1. Allgemeines zur Sequenzierstrategie	36

4.2. Bioinformatische Analysetools .....	37
4.3. Vor- und Nachteile der QIIME2 Analyse.....	37
4.4. Überblick über die Analyse des Lymphoma Datensatzes .....	39
4.4.1. Taxonomischer Vergleich mit der aktuellen Literatur .....	40
4.5. Das Darmmikrobiom der Katze im gesunden Zustand und während der Lymphomerkrankung	42
4.5.1. Firmicutes-assoziierte Bakterien.....	42
4.5.2. Bacteroidota -assoziierte Bakterien .....	43
4.5.3. Fusobacteria-assoziierte Bakterien .....	44
4.5.4. Proteobacteria-assoziierte Bakterien .....	45
4.5.5. Actinobacteria-assoziierte Bakterien .....	45
4.6. Bedeutung eines veränderten Mikrobioms der Lymphoma Katzen .....	46
4.7. Zukunft der Mikrobiomanalyse .....	48
4.8. Ausblick.....	49
<b>5. Zusammenfassung.....</b>	<b>51</b>
<b>6. Summary .....</b>	<b>52</b>
<b>7. Abkürzungsverzeichnis .....</b>	<b>53</b>
<b>8. Literaturverzeichnis .....</b>	<b>54</b>
<b>9. Abbildungs- und Tabellenverzeichnis .....</b>	<b>63</b>
<b>10. Supplement.....</b>	<b>64</b>

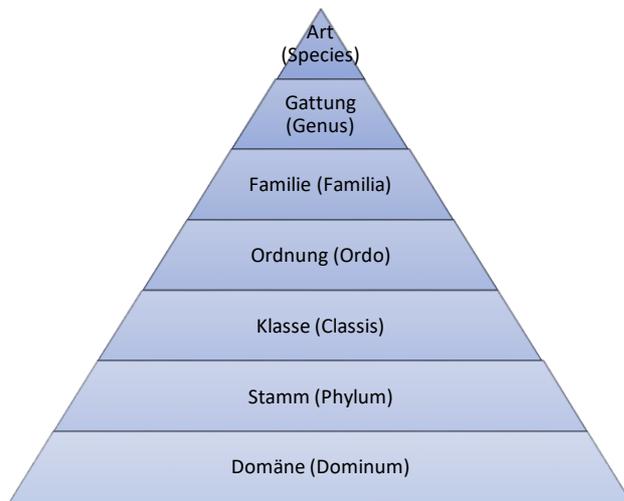
# 1. Einleitung

## 1.1. Die mikrobielle Vielfalt

Die Vielfalt der auf der Erde lebenden Mikroorganismen wird als enorm hoch eingeschätzt und ist bis heute Gegenstand aktueller Forschung. Neuen Studien zufolge ist unsere Erde Heimat für über  $10^{12}$  mikrobielle Spezies (Locey und Lennon 2016). Die Diversität umfasst ein breites Spektrum an Organismen, darunter Pilze (z.B. *Candida*), Bakterien (z.B. *Escherichia coli*), Protozoen (z.B. *Wimperntierchen*), Mikro-Algen (z.B. *Chlorellen*) und Archaeen (z.B. *Methanobrevibacter*) (Dunlap 2001). Nach wie vor werden phylogenetisch neue und (noch) nicht kultivierbare Mikroben gefunden. Es wird geschätzt, dass die nicht kultivierbaren mikrobiellen Gattungen weltweit  $7 \times 10^{29}$  Organismen umfassen könnten (Lloyd et al. 2018). Mikroben sind essenziell für den Kreislauf der Natur, da sie Produzenten in vielen Nahrungsketten sind, aber auch als Destruenten organische Stoffe zu anorganischer Matrix umwandeln. Eine Vielzahl an Mikroorganismen ist für Säugetiere von besonderer Bedeutung hinsichtlich der Erhaltung ihrer Gesundheit, darunter zum Beispiel Darmmikroben, die bei der Verdauung von Nahrung, der Bereitstellung von Vitaminen und Aminosäuren, und der Erhaltung wichtiger immunologischer Funktionen helfen (Minamoto et al. 2012, Lin et al. 2017, Wolters et al. 2019) und Hautmikroben, die die Oberfläche von Säugetieren besiedeln und so vor pathogenen Keimen schützen sowie das physiologische Hautmilieu erhalten (Sanford und Gallo 2013, Eisenstein 2020).

Molekulare Werkzeuge, die nicht auf die Kultivierung der Organismen angewiesen sind, da sie Nukleinsäuren (=genetische Information) detektieren, ermöglichen es, Mikroben qualitativ und quantitativ in *real-time* und *in vivo* zu erfassen und so neue Erkenntnisse über ihre Verwandtschaftsgrade, die Physiologie und die Genregulation unter bestimmten Umweltbedingungen zu sammeln. Taxonomische Profile und metabolische Zusammenhänge, die mit solchen molekularen Werkzeugen erforscht werden, ermöglichen ein komplexes Bild der mikrobiellen Welt, wie dies bisher nicht möglich gewesen ist (Metzker 2005).

Man klassifiziert Bakterien in dem in **Abb. 1** dargestellten taxonomischen System (Madigan 2015).



**Abb. 1** Systematische, taxonomische Einteilung der Bakterien. Ein neuer taxonomischer Name hat nur dann Gültigkeit, wenn er im "International Journal of Systematic and Evolutionary Microbiology" publiziert wurde.

## 1.2. Detektionsmethoden von Mikroben

Die Detektionsmethoden von Mikroben haben sich im Laufe der Zeit stetig weiterentwickelt. Im Folgenden wird ein Überblick über die bis heute wichtigsten Methoden gegeben.

### 1.2.1. Klassische Kultivierungsverfahren

Eine mikrobiologische Kultivierung ist ein Verfahren zur Vermehrung mikrobieller Organismen, die sich in einem Medium unter kontrollierten Laborbedingungen replizieren. Mikrobielle Kultivierung ist eine wichtige diagnostische Methode (Goldstandard in vielen Bereichen der Diagnostik und Forschung). Das Wachstum von Bakterien wird allerdings durch viele Faktoren beeinflusst, darunter Zusammensetzung der Atmosphäre, Temperatur, Inkubationszeit und Nährstoffangebot. Ein Großteil der Bakterien konnte bis heute nicht durch Kultivierungsverfahren identifiziert werden, da die Ansprüche an die Bedingungen je nach Spezies stark variieren und sehr speziell sein können (Singh et al. 2013). Bakterienkulturen erlauben auch eine Sequenzierung des Genoms, sowie den Vergleich der Transkription unter verschiedenen Bedingungen und die phylogenetische Beziehung der Stämme zueinander (Fournier et al. 2007). Eine weitere Verwendung von Reinkulturen sind Knock-out-Experimente, wobei Bakterien durch Hinzufügen oder Löschen von Genen manipuliert und transformiert werden können (Wah Tang et al. 2015). Dadurch kann die Ursache für Virulenz und Antibiotikaresistenz sowie das invasive Potenzial eines Mikroorganismus analysiert werden (Lagier et al. 2015).

Bakterien können aber auch in einen nicht kultivierbaren Zustand (VBNC; viable but non-culturable state) übergehen, dies ist eine Form der Ruhephase. Hauptmerkmal der ruhenden

Zellen ist das Herunterfahren des globalen Metabolismus und damit die Fähigkeit, sich der Nachweismöglichkeit in einer Kultur zu entziehen und stressige Umgebungen, einschließlich z.B. Angriffe des Immunsystems des Wirts oder Medikamentengaben zu tolerieren (Wong und Wang 2004). Da es nicht möglich ist in Kulturmedien Populationen von Zellen in einem Ruhezustand nachzuweisen, birgt die Verwendung solcher Standardkulturmethoden zum Nachweis pathogener Organismen immer ein Risiko und ist für die Beschreibung der Diversität der Mikroorganismen nicht optimal (Ayrapetyan und Oliver 2016). Hier ermöglichen kulturunabhängige Techniken eine schnellere und genauere Differenzierung (Austin 2017).

### 1.2.2. Auf PCR-basierende kulturunabhängige Methoden (konventionelle PCR und qPCR)

Die PCR Methode beruht auf einem Vervielfältigungsprinzip für Nukleotidstränge. Dabei können Fragmente bis zu einer Länge von ca. 3 kbp kopiert werden. Um eine PCR durchzuführen, müssen der (c)DNA verschiedene Reagenzien zugesetzt werden: ein Primerpaar, die DNA- Polymerase, Desoxyribonukleosid-Triphosphate (dNTPs) und eine Pufferlösung. Anschließend durchlaufen die Proben einen spezifischen Temperaturzyklus der 20 bis 50 Mal wiederholt wird. Ein Zyklus besteht aus jeweils drei Schritten. Erstens die Denaturierung, dabei wird die doppelsträngige DNA aufgetrennt. Zweitens das Annealing, hier lagern sich die Primer an der DNA an. Und drittens, die Elongation, dabei baut die Polymerase die komplementären dNTPs in die Einzelstränge ein. Das Ergebnis sind unzählige, doppelsträngige Kopien der DNA (Mullis et al. 1986). Da die PCR schnell eine bedeutende Rolle in der Biowissenschaft und Diagnostik darstellte, wurde diese Methode stetig im Laufe der Zeit weiterentwickelt (Zhu et al. 2020).

Die qPCR (real-time PCR) ist ein Vervielfältigungsverfahren der Nukleinsäure und basiert auf der Quantifizierung des Produkts (Amplikon). Dabei werden Endpunktmessungen nach jedem Amplifikationszyklus durchgeführt. Eine weitere Auswertung zur Bestimmung der Quantität mittels Gelelektrophorese ist somit nicht mehr notwendig (Mortarino et al. 2004). Diese Technologie ist im Gegensatz zur Kultivierung und zur normalen PCR quantitativ (Book et al. 2009). Ein Nachteil der qPCR ist allerdings die geringe Spezifität unter der Verwendung von Fluoreszenzfarbstoffen und, dass mit dem normalen qPCR Verfahren keine lebend/tot Unterscheidung gemacht werden kann. Es gibt inzwischen allerdings qPCR Assays am Markt die Reagenzien und Farbstoffe verwenden, um freie DNA zu eliminieren, sodass nur DNA auslebenden Zellen detektiert wird.

### 1.2.3. Sequenzierungsmethoden

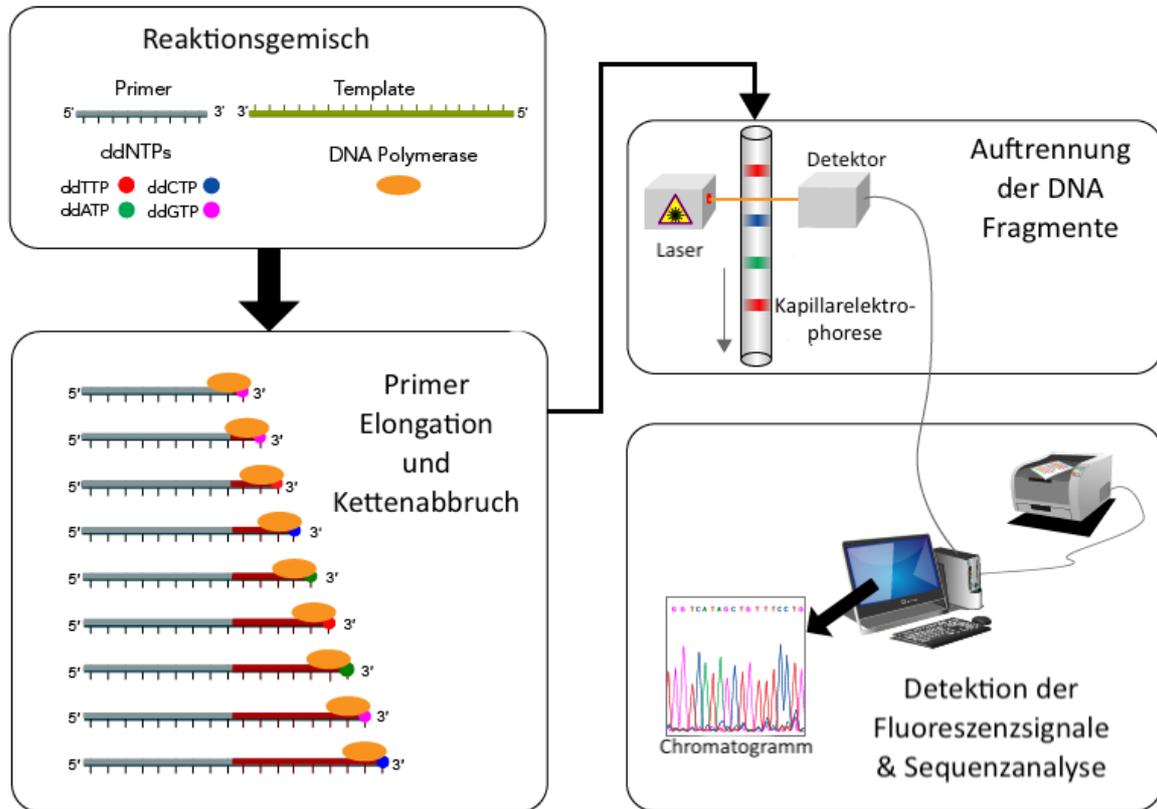
Im Jahr 1977 veröffentlichte Dr. Frederick Sanger seine Kettenabbruchmethode (Didesoxy-DNA-Sequenzierungsmethode) (Sanger et al. 1977). Durch die stetige Weiterentwicklung der Didesoxy-Methode konnten neue Generationen von Sequenzierstrategien etabliert werden, welche deutlich schneller und kostengünstiger sind (Pareek et al. 2011). Mit Hilfe dieser Technologien können Mikroben, die bis dato nicht kultivierbar waren, nachgewiesen und bestimmt werden.

Sequenzierungsmethoden werden heute in vielen Bereichen der Wissenschaft angewendet und sie werden in drei Generationen unterteilt. Die erste Generation „First-Generation-Sequencing“ (FGS) ist die Sequenzierung nach Sanger, diese bildete die Grundlage für die Entwicklung anderer Sequenzierungstechnologien welche heute als „Next-Generation-Sequencing“ (NGS) bezeichnet werden (Verma et al. 2017). Zu NGS gehören die Sequenzierung der zweiten „Second-Generation-Sequencing“ (SGS) und dritten Generation „Third-Generation-Sequencing“ (TGS). In den letzten Jahrzehnten wurden enorme Fortschritte in Bezug auf Geschwindigkeit, Leselänge und Durchsatz erzielt, zusammen mit einer starken Reduzierung der Kosten pro Basenpaar (van Dijk et al. 2014).

Heutzutage ermöglicht NGS Mikroben qualitativ und quantitativ in *real-time* und *in vivo* zu erfassen und so neue Erkenntnisse über ihre Physiologie und mikrobielle Ökologie zu sammeln. Taxonomische Profile und metabolische Zusammenhänge ermöglichen ein komplexes Bild der mikrobiellen Welt, wie dies bisher nicht möglich war (Metzker 2005).

#### 1.2.3.1. Sanger Sequenzierung

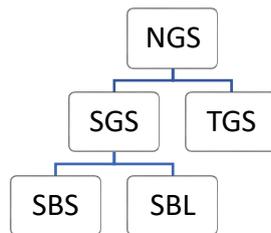
Die klassische Sanger Sequenzierung beruht auf der Kettenabbruchmethode (**Abb. 2**). Bei dieser Methode werden Didesoxyribonucleosid-Triphosphate (ddNTP) eingesetzt, welche so modifiziert sind, dass die 3'-Hydroxylgruppe fehlt. Somit können die ddNTPs keine Bindung mit dem 5'-Phosphat des nächsten dNTP eingehen. Dazu werden vier parallelen Reaktionen mit jeweils einer ddNTP-Basen (A, T, C und G) durchgeführt. Dabei wird, ähnlich wie bei einer PCR, die DNA zuerst denaturiert, sodass anschließend die Primer binden können. Mithilfe der Polymerase werden die dNTPs so lange komplementär angelagert, bis ein ddNTP eingebaut wird und die Reaktion abbricht. Dadurch entstehen unterschiedlich lange DNA-Abschnitte (Sanger et al. 1977). Für die Auswertung werden die vier Spuren auf ein Polyacrylamidgel aufgetragen oder mittels Laser detektiert (Heather und Chain 2016). Diese Technik gilt auch heute noch als Goldstandard der Sequenziermethoden (Vincent et al. 2017).



**Abb. 2** Sanger Sequenzierung modifiziert nach <https://letstalkscience.ca/educational-resources/backgrounders/sanger-sequencing> (Zugriff 11.05.2021)

### 1.2.3.2. Next-Generation Sequencing (NGS)

Zu Beginn des 21. Jahrhunderts wurde das Next-Generation Sequencing (NGS) entwickelt (Metzker 2010). Mit deren Methoden, welche sich in verschiedene Gruppen einteilen lassen (**Abb. 3**), können Wissenschaftler Tausende bis Millionen von DNA-Molekülen parallel in einem einzigen Durchlauf sequenzieren. NGS ermöglicht die Genomsequenzierung mit einer enormen Geschwindigkeit und gleichzeitig hoher Kostenersparnis (Vincent et al. 2017). So kann beispielsweise das menschliche Genom innerhalb eines Tages sequenziert werden (Behjati und Tarpey 2013).



**Abb. 3** Übersicht über die Next-Generation-Sequenzierung (NGS.) SGS (zweite Generation der Sequenzierung), TGS (dritte Generation der Sequenzierung), SBS (Sequenzierung durch Synthese) und SBL (Sequenzierung durch Hybridisierung und Ligation).

Das Grundprinzip der Sequenzierung der zweiten Generation (SGS) basiert auf vier Schritten. Der erste Schritt ist die Fragmentierung. Dabei werden DNA-Fragmente enzymatisch, mechanisch oder chemisch erzeugt. Als nächstes erfolgt die Adaption. Hier werden spezifische Adapter an die zuvor erstellten Bruchstücke gebunden. Anschließend binden diese Bruchstücke an eine stabile Oberfläche (z.B. an einen Chip) und die Amplifikation (Vervielfältigung) der DNA kann beginnen. Das Zusammenfassen von identischer DNA zu sogenannten Clustern ermöglicht eine Zeitersparnis des Sequenziervorganges. Der letzte Schritt ist die Datenanalyse. Hierbei werden die gewonnenen Daten bioinformatisch analysiert. Zu den Verfahren der SGS gehört die Sequenzierung durch Synthese (SBS) und Sequenzierungsverfahren durch Hybridisierung und Ligation (SBL).

Die Sequenzierung durch Synthese (SBS) lässt sich in drei Verfahren unterteilen. Zum einen in die Pyrosequenzierung, welche auf der Detektion von Pyrophosphat basiert. Das Pyrophosphat entsteht während der DNA-Polymerisation und die Freisetzung ist mit einem Fluoreszenzsignal gekoppelt, welches detektiert wird. Das zweite Verfahren umfasst die Sequenzierung durch reversible Termination (Illumina Sequenzierung), bei der unter Verwendung von fluoreszierenden dNTPs die Sequenzierung in Echtzeit verfolgt werden kann. Bei der Illumina-Sequenzierung, welche für diese Arbeit verwendet wurde, müssen amplifizierte „Sequencing Libraries“ hergestellt werden, bevor die Sequenzierung durchgeführt werden kann. „Libraries“ sind DNA-Fragmente mit integrierten Adaptoren (siehe 1.2.3.3. Illumina Sequenzierung).

Unter dem Verfahren Sequenzierung durch Ligation (SBL) versteht man SOLiD (Support Oligonucleotide Ligation Detection). Dessen Prinzip ist die zwei Basen-Sequenzierung. Die Sonden binden in regelmäßigen Abständen nacheinander und identifizieren dabei immer zwei Basen (Ambardar et al. 2016).

Da verschiedene Methoden existieren, orientiert man sich bei der Wahl des passenden SGS oft an der angestrebten Coverage (=Häufigkeit der Sequenzierung einer Base innerhalb einer Gesamtsequenz). Umso höher die Coverage ist, desto eindeutiger wird die Zuordnung zu Referenzsequenzen ermöglicht (Zhang et al. 2011).

Die Verfahren der zweiten Generation sind in ihrer Leselänge limitiert, weshalb die dritte Generation der NGS entwickelt wurde. Zu den Verfahren der dritten Generation (TGS) zählt Pacific Biosciences SMRT (Single Molecule Real Time), Sequencing Helicos™ Single Molecule Sequencing und Nanopore DNA Sequencing. Bei diesen Verfahren kann eine Einzelmolekülsequenzierung in Echtzeit durchgeführt werden, ohne dass die teuren und zeitaufwendigen „Libraries“ im Vorhinein generiert werden müssen.

#### *1.2.3.3. Illumina Sequenzierung*

Für diese Arbeit wurde eine Illumina MiSeq Sequenzierung der Proben durchgeführt. Im ersten Schritt wird hierfür die genomische DNA fragmentiert (~500 bp) und an beiden Enden eine Adapterligierung („Library“-Herstellung) durchgeführt. Die DNA-Fragmente binden dann zufällig über die Adapter an die Oberfläche einer Illumina Flow Cell (spezieller Glasträger), welche mit komplementären Adaptermolekülen versehen ist. Dadurch, dass beide Enden an einem freien Adapter auf der stationären Phase binden, entsteht eine DNA-Brücke. Die Polymerase beginnt den komplementären Strang zu synthetisieren und die sogenannte Brücken-PCR läuft ab. Es entstehen eine Vielzahl an DNA-Fragmenten mit identischer Sequenz. Diese Kopien werden als Cluster bezeichnet. Im letzten Schritt baut die Polymerase die zuvor mit unterschiedlichen fluoreszierenden Farbstoffen gekennzeichneten dNTPs ein <https://www.illumina.com/science/technology/next-generation-sequencing/illumina-sequencing-history.html> (Zugriff 11.05.2021).

Die Hochdurchsatzsequenzierung des 16S rRNA Gens auf der Illumina-Plattform wird zur Beurteilung der mikrobiellen Vielfalt in vielen Umweltproben verwendet (Pichler et al. 2018). Illumina-Sequenziermaschinen bieten unterschiedliche Durchsatzraten an. Der MiniSeq beinhaltet ein breites Spektrum an Möglichkeiten, bei gleichzeitig geringem Platzverbrauch und einem verhältnismäßig geringen Anschaffungspreis. Dieses Produkt soll auch den kleinsten Labors Zugang zu Illumina ermöglichen. 2011 veröffentlichte Illumina mit MiSeq ein Instrument für den schnellen Durchlauf mit niedrigerem Durchsatz, das sich an kleinere Labors und für das Erstellen von klinischen Diagnosen richtet (Quail et al. 2012). Ein weiteres Gerät ist das NextSeq, welches bewährte Technologien, unter anderem Next-Generation-Sequencing und hochauflösendes Array Scanning, kombiniert. Mit dem HiSeq 2000 hat die Firma den Standard für die parallele Sequenzierung mit hohem Durchsatz gesetzt <https://www.illumina.com/systems/sequencing-platforms.html> (Zugriff 11.05.2021). In **Tab. 1** sind die Eigenschaften der unterschiedlichen Systeme von Illumina dargestellt.



	<b>MiniSeq</b>	<b>MiSeq</b>	<b>NextSeq 1000&amp;2000</b>	<b>HiSeq 2500</b>
Output	7.5 Gb	0.3–15 Gb	40-330 Gb	9 Gb - 1 Tb
Laufzeit pro Lauf	4-24 h	5–55 h	11-48 h	7 h - 6 d
Sequenzen pro Lauf	25 Mio	1–25 Mio	0.4–1.1 Mill	300 Mio- 4 Mill
Max. Sequenzlänge	2 x 150 bp	2 x 300 bp	2 x 150 bp	2 x 250 bp
Proben pro Durchlauf	50	bis zu 384	30-75	bis zu 260

**Tab. 1** Vergleich der verschiedenen Illumina Sequenziermaschinen modifiziert nach <https://www.illumina.com/systems/sequencing-platforms/comparison-tool.html#/research-use-only/microbiology/small-whole-genome-sequencing> (Zugriff 11.05.2021). Mio= Millionen, Mill= Milliarden; h=Stunden, d= Tage bp= base pairs, Gb= giga base pairs, Tb= tera base pairs.

### 1.3. Bioinformatische Analyse von Hochdurchsatzsequenzierungen

Die Bioinformatik ist ein interdisziplinäres Gebiet, welches die Informatik und Biostatistik mit biologischen Wissenschaften wie Genetik, Genomik und Physiologie verbindet. Ziele der Bioinformatik sind die Computer gestützte Analyse, Verwaltung, Speicherung und Interpretation von biologischen Daten (Moore 2007). Die Entwicklung von Tools für die bioinformatische Analyse ist notwendig, um die molekular biologischen relevanten Informationen aus den Daten der Hochdurchsatzsequenzierung zu gewinnen, auszuwerten und korrekt interpretieren zu können. Die Hochdurchsatzsequenzierung ermöglicht es uns, in relativ kurzer Zeit eine Momentaufnahme einer mikrobiellen Gemeinschaft zu dem Zeitpunkt der Probenentnahme zu erhalten (Escobar-Zepeda et al. 2015).

Haynes et al. nennen acht Schritte zum Aufbau eines effektiven Workflows (Haines und Pericak-Vance 2007). Der erste Schritt ist die Entwicklung des Versuchsaufbaus und das Herausarbeiten der spezifischen Anforderungen an die Analyse. Im zweiten Schritt werden die Rohdaten aus dem Labor eingespielt. Der dritte Schritt ist die Datenspeicherung, um anschließend den Zusammenhang der Ergebnisse nachvollziehen zu können. Im vierten Schritt werden die Anforderungen an die Hardware- und Software bestimmt. Dabei ist es wichtig zu wissen, welche Datenmenge gespeichert werden muss, wie schnell auf diese zugegriffen werden soll und welches Betriebssystem verwendet wird. Der fünfte Schritt umfasst das sogenannte Implementieren. Es besteht daraus, die Daten in ein funktionsfähiges Analyseprogramm zu integrieren. Der sechste Schritt ist die Auswahl einer passenden

Benutzeroberfläche. Im siebten Schritt werden die Sicherheitsanforderungen ermittelt. Der letzte und achte Schritt besteht darin, die Softwaretools auszuwählen und anzuwenden, um die gewonnenen Daten aus der Analyse und die Zusammenfassung darzustellen. Um diese Prozesse zu erleichtern, gibt es immer mehr öffentliche Datenbanken, die für die Allgemeinheit zugänglich sind und somit die Analyse von Daten enorm erleichtern (Haines und Pericak-Vance 2007).

Durch die Weiterentwicklung von NGS können riesige Datensätze generiert werden, die eine enorme Rechenleistung für die Analyse benötigen. Je größer der generierte Datensatz ist, desto höhere Rechenressourcen und komplexere Bioinformatik Analysen sind erforderlich. Nicht nur die Datenverarbeitung, sondern auch die Speicherung und Archivierung erfordert enorme Serverkapazität (Logares et al. 2012). Die Tools zur Analyse sind nach wie vor rar, sehr spezifisch und komplex (Caporaso et al. 2010). Die Analyseparameter in den Programmen sind oft nicht variabel genug für die Ansprüche der Datenauswertung und die Übertragung der großen Datenmengen stellt weitere Herausforderungen für die Tools dar. Ziel ist es, klare, einheitliche, bioinformatische Verfahren zu entwickeln, um einen reibungslosen und reproduzierbaren Ablauf der Datenverarbeitung zu gewährleisten. Auf diese Weise kann die Qualität und die Vertrauenswürdigkeit der Ergebnisse gesichert werden (Funari und Canosa 2014).

### 1.3.1. Vergleich von bioinformatischen Plattformen

Die 16S rRNA Gen Sequenzierung ist die am häufigsten genutzte Methode zur Erstellung eines mikrobiellen Profils (Douglas et al. 2020). Die verschiedenen Analysetools die derzeit zugänglich sind weisen unterschiedliche Schwerpunkte in der Datenanalyse auf. Es ist daher wichtig, im Voraus die Fragestellung des Versuchs zu definieren und mit den Besonderheiten der Programme vertraut zu sein. Nur so kann die Reproduzierbarkeit und Vergleichbarkeit von Studien erhöht und aussagekräftige Ergebnisse gewonnen werden (Allali et al. 2017).

„End-to-end“ Analysetools, die für Rohdaten bis hin zur finalen Analyse genutzt werden können:

QIIME2: In QIIME 2 (Quantitative Insights Into Microbial Ecology 2; open source software) erfolgt die Sequenzqualitätskontrolle über verschiedene Tools, die in Form von Plugins integriert sind. Die Besonderheiten von QIIME 2 gegenüber anderen Analyseprogrammen sind interaktive Visualisierungstools für die mikrobielle Gemeinschaft, welche ohne zusätzliche Installationen genutzt werden können. Des Weiteren werden alle Analyseschritte automatisch als Zwischenergebnisse gespeichert, um die Rückverfolgbarkeit der Daten zu gewährleisten und somit eine Reproduzierbarkeit der Analyse zu ermöglichen (Bolyen et al. 2019a). Eine

Standard-QIIME-Analyse beginnt mit der Einspielung von Sequenzdaten von einer oder mehreren Sequenzierungsplattformen wie Sanger, Roche / 454 oder Illumina. Es folgt die Eingabe einer Reihe von Befehlen in ein Terminalfenster, um anschließend die Textausgabe und graphische Auswertung anzuzeigen. Die Hauptausgabe der Analyse ist eine sogenannte Feature-Tabelle, welche die Häufigkeit aller OTUs (operational taxonomic units) oder ASVs (amplicon sequencing variants) in jeder Stichprobe beschreibt (Kuczynski et al. 2011). OTUs/ASVs sind die Basis für phylogenetische Bäume sowie für die phylogenetische und taxon-basierende Analyse der Diversität innerhalb und zwischen Proben (Caporaso et al. 2010). QIIME unterstützt eine breite Palette von Analysen und Visualisierungen der mikrobiellen Gemeinschaft. Des Weiteren nutzt dieses Programm in großem Umfang Unit-Tests (Komponententests/Modultests), um die Signifikanz der Ergebnisse zu prüfen. Viele Mikrobiota Studien wurden mit QIIME™ analysiert und durch direkte Zusammenarbeit von Entwickler und Benutzer, z.B. über das Online-Support-Forum, permanent verbessert, sodass die Plattform einem immer größer werdenden Feld an Mikrobiomforschern, sowohl in der Wissenschaft, als auch in der Industrie, von Nutzen ist (Bolyen et al. 2019b).

Mothur: Mothur ist auch eine beliebte Open-Source-Software für die Analyse von Amplicon Sequenzier-Datensätzen. Das Besondere an diesem Tool ist, dass es vom Prozessieren der Rohsequenzen und Visualisierung bis hin zur Alpha- und Beta-Diversität alles in einem Programm vereint. Somit ermöglicht es dem Benutzer, zur Auswertung nur ein einziges Tool zu nutzen. Mothur ist ein effizientes, flexibles und leistungsstarkes Open-Software-Tool (Schloss et al. 2009). Vergleicht man QIIME mit Mothur, zeigen sich keine statistisch signifikanten Unterschiede bei der Schätzung der relativen Häufigkeit auf Gattungsebene und der Anzahl an Taxa, unter der Verwendung von SILVA als Referenzdatenbank. Nutzt man jedoch Green Genes als Referenzdatenbank, zeigt Mothur eine höhere Diversität als QIIME. Des Weiteren gibt es zwischen Mothur und QIIME Abweichungen bei der Berechnung der Beta-Diversität. Diese Abweichungen werden durch die unterschiedlichen Schätzalgorithmen der selten vorkommenden Mikroorganismen der beiden Analysetools ausgelöst (López-García et al. 2018).

DADA2: DADA 2 (Divisive Amplicon Denoising Algorithm 2) ist ein Tool zur Analyse der mikrobiellen Gemeinschaft und kann nur für Daten aus der Illumina Sequenzierung angewendet werden. Es ist für alle Gensequenzen geeignet und auch in QIIME 2 integriert. Fehler in den Sequenzen werden mithilfe von Qualitätsfilterung und Bildung von OTUs korrigiert. Innerhalb der OTUs werden ähnliche Sequenzen zusammengefasst, wodurch eine Fehlerreduktion erfolgt und die Fehlinterpretationen der Sequenzen vorgebeugt wird (Callahan et al. 2016). Anstelle von OTUs können auch ASVs genutzt werden. Durch ASVs können Unterschiede von einem Nukleotid in den Sequenzdaten analysiert werden. Durch die Verwendung von ASVs ist eine taxonomische Zuordnung genauer als durch eine OTU

Klassifizierung und der direkte Vergleich von Daten aus unterschiedlichen Studien ist so erstmals möglich <https://benijneb.github.io/dada2/index.html> (Zugriff am 25.09.21).

**MICCA:** MICCA (MICRobial Community Analyse) ist eine öffentlich zugängliche Softwarepipeline zur Verarbeitung von Single-End Sequenzen, welche durch Roche 454 oder Illumina Sequenzierung gewonnen werden. Mit diesem Tool können Qualitätsfilterung, Clustern von OTUs und taxonomische Klassifizierung kombiniert werden und mikrobielle Gemeinschaften analysiert werden. Ein neuer Clustering Algorithmus, welcher speziell für OTUs entwickelt wurde, liefert bei geringer Sequenztiefe genaue Schätzungen über die Anzahl an OTUs in der Analyse. Im Vergleich dazu erfordern andere Tools (QIIME und Mothur) eine bessere Sequenztiefe. MICCA ist Python-basiert und kann auf allen gängigen Computeroberflächen genutzt werden, zudem ist es speichereffizient und einfach zu bedienen (Albanese et al. 2015) <https://compmetagen.github.io/micca/> (Zugriff 26.09.21).

Da die Nachfrage an Analysetools stetig steigt, werden immer neue end-to-end Tools publiziert, unter anderem **rANOMALY:** rANOMALY (Amplicon workflow for Microbial community AnaLYsis) ist ein auf R-Paket basierendes Analysetool. Ziel ist es, durch die Verwendung eine Standardisierung und Automatisierung zu ermöglichen. Dadurch steigt die Reproduzierbarkeit von Ergebnissen und die Vergleichbarkeit verschiedenen Studien miteinander wird erhöht (Theil und Rifa 2021).

**Microeco:** ist ebenfalls ein auf einem R-Paket basierendes Tool zur Analyse von mikrobiellen Gemeinschaften. Es kombiniert eine Reihe von häufig verwendeten Analyseschritten, wie zum Beispiel die Alpha-, und Beta-Diversitätsanalyse. Microeco stellt für den Benutzer ein schnelles, flexibles und leistungsstarkes Analysetool dar (Liu et al. 2021).

**gDAT:** gDAT (graphical Downstream Analysis Tool) ist eine Python basierte Pipeline. Durch die Kombination der gängigsten Kommandozeilen-Analysetools und anschließende Integration in eine benutzerfreundliche graphische Oberfläche, ist dieses Programm anpassungsfähig und zukunftsorientiert (Vasar et al. 2021).

Nicht nur die Auswahl des Analysetools, sondern auch die der Referenzdatenbanken, spielen eine große Rolle bei der Datenanalyse. SILVA wird im Gegensatz zu Green Genes regelmäßig aktualisiert, weshalb SILVA den bevorzugten Referenzdatensatz zur Klassifizierung der Daten darstellt (Allali et al. 2017, López-García et al. 2018).

## 1.4. Darmmikrobiota

Molekular-phylogenetische Studien haben gezeigt, dass das intestinale Mikrobiom von Säugetieren sehr unterschiedlich ist und mehrere hundert bis über tausend bakterielle Spezies beherbergt (Minamoto et al. 2012b). Die mikrobielle Gesamtheit im Darm von Säugetieren wird auf  $10^{12}$  bis  $10^{14}$  Organismen geschätzt. Dies entspricht ca. der zehnfachen Menge der

Wirtszellen (Suchodolski et al. 2009, Handl et al. 2011). Neben der vertikalen Übertragung von Mikrobiota vom Muttertier auf das Jungtier wird der gegenseitige, permanente Mikrobenaustausch mit der Umwelt als enormer Vorteil des Gruppenlebens angesehen (Lombardo 2008). Von 55 bekannten bakteriellen Phyla dominieren nur drei den Säugetier-GIT: *Bacteroidota*, *Firmicutes* und *Proteobacteria*. Daher ist es wahrscheinlich, dass sich Wirtsorganismen mit ihren GIT-assoziierten Mikroben co-evolutionär entwickelt haben, um von der Symbiose die zur gastrointestinalen Homöostase beiträgt, zu profitieren (Bäckhed et al. 2005).

Die Gesamtheit aller mikrobiellen Gene im GIT (Gastrointestinaltrakt) wird als Darmmikrobiom bezeichnet. Dies spielt für die Gesundheit des Wirts eine entscheidende Rolle. Durch die ansässigen Mikroorganismen kann das Immunsystem stimuliert und Krankheitserreger abgewehrt werden (Minamoto et al. 2012b). Des Weiteren unterstützen Mikrobiota die Verdauung und bieten somit einen ernährungsphysiologischen Nutzen für Enterozyten, z.B. durch die Produktion kurzkettiger Fettsäuren (Tremaroli und Bäckhed 2012). Ein intaktes Mikrobiom bietet eine Vielzahl von gesundheitlichen Vorteilen, zum Beispiel dient es als Barriere gegen das Eindringen von Krankheitserregern (Neish 2009). Verschiedene exogene Faktoren, wie Ernährung oder Haltung, beeinflussen die Zusammensetzung der Mikrobiota, dennoch ist diese gegenüber den meisten Umwelteinflüssen stabil und kehrt nach Behandlungen meist zu der ursprünglichen Zusammensetzung zurück (Lyu et al. 2020). Frühere Studien haben gezeigt, dass die Immunität des Wirts gegenüber Krankheiten in einer engen Wechselwirkung mit einem intakten Mikrobiom des GIT steht. Eine mikrobielle Dysbiose kann mit unterschiedlichen gastrointestinalen Krankheiten und extraintestinalen Störungen bei Säugetieren in Verbindung gebracht werden (Suchodolski 2011a, Lyu et al. 2020).

In dieser Arbeit wurden Mikrobiota von an Lymphomen erkrankten Katzen mit einer gesunden Kontrollgruppe anhand von Kotproben verglichen. Bei gesunden Katzen nimmt die Anzahl der Mikroorganismen im GI-Trakt vom Magen zum Dickdarm zu (Handl et al. 2011). In den proximalen Teilen des GI-Trakts wird eine gleichmäßigere Verteilung der aeroben und anaeroben Bakterien beobachtet, während in den distalen Teilen des GI-Trakts hauptsächlich anaerobe Bakteriengruppen überwiegen (Lyu et al. 2020). Die dominierenden Bakteriengruppen ähneln denen, die bei anderen Säugetieren gefunden wurden, wobei die Phyla *Firmicutes*, *Bacteroidota*, *Actinobacteria* und *Proteobacteria* mehr als 99% der Darmmikrobiota ausmachen (Minamoto et al. 2012b, Deng und Swanson 2015). Dabei sind *Firmicutes* und *Bacteroidota* die abundantesten Phyla im fäkalen Mikrobiom der Katze (Moon et al. 2018). Die genaue Zusammensetzung des bakteriellen Mikrobioms im Katzendarm variiert zwischen verschiedenen Studien, es werden aber in den meisten Studien folgende dominante Phyla beschrieben: *Firmicutes*, *Proteobacteria*, *Bacteroidota*, *Fusobacteria* und

*Actinobacteria* (Ritchie et al. 2008, Suchodolski 2011a, Tun et al. 2012). Abweichungen in den relativen Abundanzen dieser Phyla, die man findet, wenn man die Proben dieser drei Arbeiten vergleicht, könnten einerseits mit unterschiedlichen Umwelteinflüssen, Fütterungsunterschieden und/oder unterschiedlicher Haltung der Tiere erklärt werden. Zusätzlich könnten nicht standardisierte Probeentnahmen und Analysen verantwortlich für Abweichungen der Ergebnisse sein (Lyu et al. 2020).

## 1.5. Hypothese und Ziel dieser Arbeit

Hypothese dieser Arbeit: An Lymphomen erkrankte Katzen zeigen einen nachweisbaren Mikrobiota Shift (Dysbiose) im Kot und unterscheiden sich im Mikrobiota Profil signifikant von gesunden Katzen.

Ziel der Arbeit: Analyse des NGS Datensatzes mit QIIME2 und biologische Interpretation der Ergebnisse.

## 2. Material und Methoden

Kotproben von 33 Katzen wurden aus der Vetbiobank entnommen. Die DNA aus dem Kot wurde mit dem PowerSoil DNA Extraktionskit gewonnen und die Qualität und Quantität der DNA vermessen (**Tab. 3**). Nach internen PCRs, die das Vorhandensein von genügend Biomasse bestätigten, wurde eine MiSeq Illumina Hochdurchsatzsequenzierung durchgeführt und der Datensatz mit QIIME2 ausgewertet.

### 2.1. Probengewinnung

Die Gewinnung der Kotproben wurde von Mitarbeiter/innen (unter der Leitung von Priv.-Doz. Dr. Birgitt Wolfesberger) der Klinischen Abteilung für Interne Medizin Kleintiere (Veterinärmedizinische Universität Wien, Österreich) ausgeführt. Hierbei wurden 33 Kotproben gewonnen. Davon sind 23 Proben von Lymphom erkrankten und 10 von gesunden Katzen. Für diese Arbeit wurden die gesammelten Kotproben der erkrankten Tiere aus der VetBiobank (Technologieplattform VetCore) entnommen.

### 2.2. DNA-Extraktion

Die Isolierung von DNA erfolgte aus 250 mg Kot mit Hilfe des kommerziellen Kits DNeasy PowerSoil Pro (Qiagen, Hilden, Deutschland) nach Vorgaben des Herstellers (Protokoll, „DNeasy PowerSoil Pro Kit Handbook“

<https://www.qiagen.com/us/resources/resourcedetail?id=9bb59b74-e493-4aeb-b6c1-f660852e8d97&lang=en> (Zugriff 14.05.21)). Eine Leerkontrolle wurde inkludiert.

### 2.3. DNA-Vermessung

Die Konzentrationsbestimmung der doppelsträngigen DNA erfolgte fluorometrisch mit einem Qubit 2.0 Fluorometer (Invitrogen, Vienna, Austria). Die verwendete QubitDye dsDNA BR Reagenz bindet dabei selektiv an doppelsträngige DNA. Über den Fluoreszenzfarbstoff konnte die DNA-Konzentration der Probe bestimmt werden. Die Durchführung der DNA-Bestimmung erfolgte wie im Qubit® dsDNA BR Assay Kits Protokoll beschrieben (“Qubit 2.0 Fluorometer,” <https://www.thermofisher.com/document-connect/document->

[connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2FQubit\\_dsDNA\\_BR\\_Assay\\_UG.pdf&title=VXNlciBHdWlkZTogUXViaXQgZHNETkEgQllgQXNzYXkgS2l0cw==](https://assets.thermofisher.com/FTFS-Assets/FLSG/manuals/Qubit_dsDNA_BR_Assay_UG.pdf&title=VXNlciBHdWlkZTogUXViaXQgZHNETkEgQllgQXNzYXkgS2l0cw==) (Zugriff 15.05.21)).

## 2.4. Sequenzierung

Das Vorhandensein von genügend Biomasse wurde zuerst mit Hilfe von der 16S rRNA Gen PCR mit den Primern (27F (5'- AGAGTTTGATCMTGGCTCAG-3') and 1492R (5'- GGYTACCTTGTTACGACTT-3') überprüft.

Für die Hochdurchsatzsequenzierung wurden Libraries für die Amplifikation des 16S rRNA-Gens (V3/4-Region) basierend auf den Empfehlungen zur Vorbereitung der 16S Metagenomic Sequencing Library von Illumina hergestellt. Primer 341F (5'-CCTACGGGNGGCWGCAG-3') und 805R (5'-GACTACHVGGGTATCTAATCC-3') (Klindworth et al. 2013) wurden zusammen mit Illumina-Adaptersequenzen (5' CGTCGGCAGCGTCAGATGTGTATAAGAGACAG bzw. 5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG) für die Amplifikation verwendet. Die Libraries wurden durch Ligieren von Sequenzierungsadaptern und Indizes auf gereinigte PCR-Produkte unter Verwendung des Nextera XT Sample Preparation Kit hergestellt (Illumina). Äquimolare Mengen von jedem der aufgereinigten Amplikons wurden gepoolt und auf einem Illumina MiSeq Sequencer mit einem 300-bp „paired-end“ Leseprotokoll sequenziert. Die Generierung und Sequenzierung der 16S-rRNA-Gen-Amplikon Libraries wurde in der NGS-Einheit der Vienna Biocenter Core Facilities ([www.vbcf.ac.at](http://www.vbcf.ac.at)) durchgeführt.

## 2.5. Bioinformatische Analyse

Es wurde ein Protokoll für die bioinformatische Analyse erstellt:

### 1. Starten der virtuellen Computeroberfläche (Virtual Box) im Conda Environment

Nutzen von der Version: qiime2-2019.7

```
conda activate qiime2-2019.7
```

### 2. Erstellen eines neuen Ordners mit dem Namen: trimmomatic

```
mkdir trimmomatic
```

### 3. Vorbereiten der Einspielung der Rohdaten – Anlegen eines Manifestfiles

Ein Textdokument wurde erstellt (**Supplement Tab. 1**), wo die fastq Namen der Proben gelistet sind. Diese wurden verknüpft mit dem jeweiligen Dateipfad, also der Information wo die Rohdaten zu finden sind.

#### 4. Entfernen der Adapter und Primer

Hierfür wurden die Sequenzen mit einer Datenbank verglichen, die eine Vielzahl an Adaptor- und Primervarianten beinhaltet (trimmomatic).

```
for i in $(ls raw-fastq/*R1* | sed "s/.*V//" | sed "s/_R1.fastq.gz//"); do
  trimmomatic PE -threads 22 -phred33 raw-fastq/${i}_R1.fastq.gz raw-fastq/${i}_R2.fastq.gz
  trimmomatic/${i}_R1.trimmomatic.fastq.gz /dev/null trimmomatic/${i}_R2.trimmomatic.fastq.gz
  /dev/null ILLUMINACLIP:/Daten2/Peffi/Cat_Lymphoma/QIIME2_conda/adapters.fasta:1:30:11
```

#### 5. Anlegen eines neuen Ordners (02-qiime)

```
mkdir 02-qiime
```

#### 6. Import von dem Manifestfile in QIIME und Zusammenfassung der Anzahl und Qualität der Sequenzen

Der Output dieses Kommandos ermöglicht es, eine interaktive Zusammenfassung der Vorwärts- und Rückwärtssequenzen zu bekommen. Diese Zusammenfassung enthält Informationen über die Qualität des Sequenzierungslaufs und die Anzahl der pro Probe erhaltenen Sequenzen.

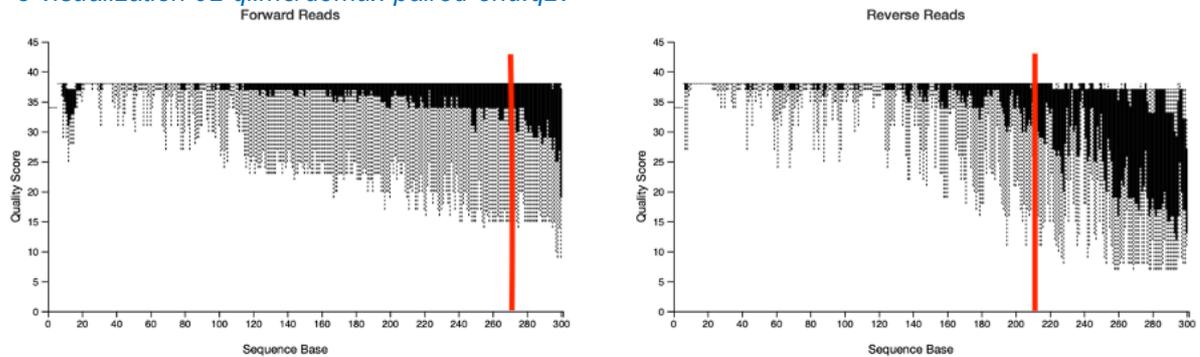
```
qiime tools import \
  --type 'SampleData[PairedEndSequencesWithQuality]' \
  --input-path manifest_file.txt \
  --input-format PairedEndFastqManifestPhred33V2 \
  --output-path 02-qiime/demux-paired-end.qza
```

#### 7. Visualisierung der Qualität der Sequenzen

Visuelle Kontrolle der Qualität von 5' und 3' Enden über den *QIIME2 Viewer* (<https://view.qiime2.org/>) aus der Datei *-demux-paired-end.qzv*. Mithilfe eines interaktiven Diagramms (**Abb. 4**) kann man feststellen, ob an einem bestimmten Punkt ein Qualitätsverlust im base calling, also der Zuordnung der Basen, auftritt. Dies kann bei der Auswahl der Kürzungs- und Trimmparameter im nächsten Schritt hilfreich sein (Estaki et al. 2020). Bei Illumina Sequenzierungen ist bekannt, dass die Qualität an den höheren Basenpositionen abnimmt. An dem Punkt des Beginns der Qualitätsabnahme sollte der Cutoff gelegt werden und die Sequenzen sollten hier im nächsten Schritt abgeschnitten werden. Der Cutoff wird für Vorwärts- und Rückwärtssequenzlesevorgänge separat bestimmt (Hall und Beiko 2018). Bei dem dieser Arbeit zugrunde liegenden Datensatz

wurde der Cutoff, basierend auf **Abb. 4**, bei den Vorwärtssequenzen bei 270 bp und bei den Reverssequenzen bei 210 bp gelegt.

```
qiime demux summarize \
--i-data 02-qiime/demux-paired-end.qza \
--o-visualization 02-qiime/demux-paired-end.qzv
```



**Abb. 4** Ausschnitt aus dem QIIME2 Viewer, Qualitätsbestimmung der Vorwärts- und Rückwärtssequenzen und visuelle Legung des in rot gekennzeichneten Cutoffs an der Stelle, wo die Qualität der Basenbestimmung abnimmt.

#### 8. Denoising: Beschneiden der Sequenzen beim Cutoff, Qualitätsfilterung der Sequenzen und Filtern von Chimären

Für die Eingabe zum Beschneiden der Sequenzen wurden zwei Größen benötigt. Zum einen `--p-trunc-len-f`, welcher die Position angibt, an der die Vorwärtssequenz abgeschnitten wird und zum anderen `--p-trunc-len-r`, an der die Rückwärtssequenz beschnitten werden soll. Die Primer wurden zusätzlich mit folgender Eingabe aus allen Sequenzen entfernt: `--p-trim-left-f` und `--p-trim-left-r`, dies ist auf die Länge der Primersequenzen (15 bp) eingestellt. Der Output ist zum einen eine Tabellendatei, welche die beschnittenen Sequenzen mit Rückverfolgbarkeit zu den Ursprungssequenzen enthält. Zum anderen ist er eine repräsentative Sequenzdatei, welche nur die qualitäts-geprüften Sequenzen enthält.

```
qiime dada2 denoise-paired \
--i-demultiplexed-seqs 02-qiime/demux-paired-end.qza \
--p-trim-left-f 15 \
--p-trim-left-r 15 \
--p-trunc-len-f 270 \
--p-trunc-len-r 210 \
--p-max-ee-f 2 --p-max-ee-r 2 \
--o-representative-sequences 02-qiime/rep-seqs.qza \
--o-table 02-qiime/table.qza \
--o-denoising-stats 02-qiime/stats-dada2.qza \
--p-n-threads 22
```

#### 9. Erstellen und Kontrollieren des Metadatenfiles

Alle Normen und Eigenschaften des Manifestfiles wurden von QIIME2 in das Metadatenfile (**Supplement Tab. 3**) übernommen. Metadaten sind Informationen zu den Proben, z.B.

Behandlung, Probenart oder Gesundheitszustand. Diese Informationen sollten in einer einzelnen Tabelle enthalten sein, die die Beispiele (Proben) als Zeilen und die Variablen (Typ, Gruppe, Behandlung etc.) als Spalten beinhaltet. QIIME2 unterstützt derzeit kategoriale und numerische Metadatentypen. Für das Metadatenfile wurden folgende Informationen benötigt und zusammengefasst: die Proben-ID, Barcode-Sequenz, Probenart (Kot), Gruppe (Gesund/Lymphoma) und „Treatment“ (behandelt/keine Behandlung). Somit besteht diese Datei aus vier kategorialen Säulen (**Tab. 2**).

*qiime tools inspect-metadata sample-metadata.tsv*

COLUMN NAME	TYPE
barcode-sequence	categorical
SampleType	categorical
Group	categorical
Treatment	categorical
IDS:	34
COLUMNS:	4

**Tab. 2** Zusammenfassung des Metadatenfiles, bestehend aus 34 Zeilen und 4 Spalten.

#### 10. Erstellen der ASV (Amplicon Sequence Variant) Tabelle

Mit dem Befehl *qiime feature-table summarize* wurde eine Visualisierung (**Abb. 5**) erstellt, welche detaillierte Informationen zu der Anzahl der entrauschten Sequenzen und der Anzahl der Sequenzen pro Probe enthält. Die visuelle Begutachtung des Outputs erfolgt auch hier über den QIIME2 Viewer. In **Abb. 5** sieht man einen Screenshot des QIIME2 Viewers, es wird gezeigt wie viele Sequenzen pro Probe vorhanden sind und wie viele ASVs detektiert wurden.

*qiime feature-table summarize*  
*--i-table 02-qiime/table.qza*  
*--o-visualization 02-qiime/table.qzv*  
*--m-sample-metadata-file sample-metadata.tsv*

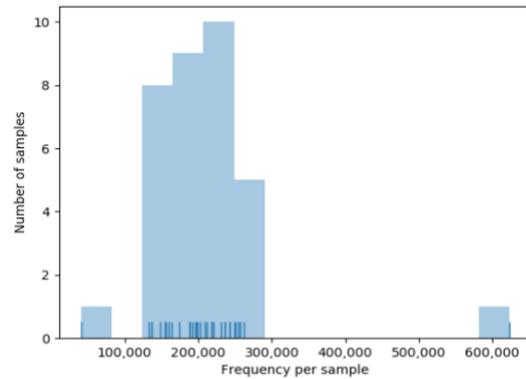
### Table summary

Metric	Sample
Number of samples	34
Number of features	3,005
Total frequency	7,122,273

### Frequency per sample

	Frequency
Minimum frequency	40,459.0
1st quartile	167,047.25
Median frequency	200,585.0
3rd quartile	236,536.25
Maximum frequency	623,474.0
Mean frequency	209,478.61764705883

Frequency per sample detail ([csv](#) | [html](#))

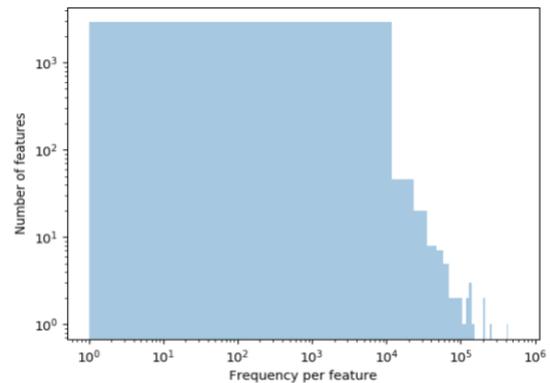


[Download as PDF](#)

### Frequency per feature

	Frequency
Minimum frequency	1.0
1st quartile	4.0
Median frequency	33.0
3rd quartile	327.0
Maximum frequency	585,792.0
Mean frequency	2,370.140765391015

Frequency per feature detail ([csv](#) | [html](#))



**Abb. 5** Ausschnitt aus dem QIIME2 Viewer, der die Ergebnisse der ASV Feature-Table zeigt. Es ist der Überblick über die Anzahl der Sequenzen und der ASVs (features) inklusive deren Verteilung gezeigt.

Im nächsten Schritt wurde eine tabellarische Ansicht der ASV Feature-ID mit Sequenzzuordnung erstellt:

```
qiime feature-table tabulate-seqs
--i-data 02-qiime/rep-seqs.qza
--o-visualization 02-qiime/rep-seqs.qzv
```

In **Abb. 6** ist ein Ausschnitt aus der ASV-Tabelle dargestellt. In dieser Tabelle sind die Feature-IDs (ASV-IDs) den dazugehörigen Sequenzen, mit Link hinterlegt, zugeordnet. Über den Link kann jede Sequenz auf der NCBI nt-Datenbank (<https://www.ncbi.nlm.nih.gov/nucleotide/>) geladen werden. Die NCBI-BLAST-Links sind nützlich, um die taxonomische Ähnlichkeit zur Referenzdatenbank zu prüfen (Estaki et al. 2020).

Feature ID	Length	Sequence
d059cf6b6634be17f15a238135743a8e	409	AGTGGGGAATATTGCACAAATGGGCGCAAGCCTGATGCAGCAACGCCCGTGAGCGATGAAGGCCCTCGGGTCGTAAGCTCTGTCTCAAGGAAGATAATGACGGT
47959156085c7e6b7bdb022c16c749f5	413	AGTGGGGAATATTGCACAAATGGACCAAAAAGTCTGATCCAGCAATCTGTGTGCACGATGACGGTCTTAGGATTGTAAGTCTTTCAATCGGGAAAAAGAAAGTGA
0f2f963e24b81d523800c9664cb8fc21	430	AGTGAGGAATATTGGTCAATGGGCGCAGGCCCTGAACCAGCAAGTAGCGTGAAGGATGACTGCCCTATGGGTTGTAAGCTCTTTTATAGGGAATAAGTGTCC
e9f922b0323f3833ef6e7dae13f71d54	430	AGTGAGGAATATTGGTCAATGGGCGAGAGTCTGAACCAGCCAAAGTAGCGTGAAGGATGACTGCCCTATGGGTTGTAAGCTCTTTTATAGGGAATAAGTGTCC
e2985dada1c89944d7081554b1b0ef1	410	AGTGGGGAATATTGCACAAATGGGGAAACCCCTGATGCAGCAACGCCCGTGAGTGAAGGTTTTTCGGATCGTAAGCTCTGTCTTTGGGGAAGATAATGACGGT
562f8e5fe4e773855042ddd638cddf26	430	AGTGAGGAATATTGGTCAATGGGCGCTAGCCTGAACCAGCAAGTAGCGTGAAGGATGAAGGCCCTATGGGTCGTAAGCTCTTTTATATAAGAATAAGTGCAGT
10a52f96ba29e1ee2c6893459e56a7b2	430	AGTGAGGAATATTGGTCAATGGGCGAGAGTCTGAACCAGCCAAAGTAGCGTGAAGGATGACTGCCCTATGGGTTGTAAGCTCTTTTATACGGGAATAAGTGTAGGC
8cbd6dabb13df66486da45568f260de6	409	AGTGGGGAATATTGCACAAATGGGCGAAGCCTGATGCAGCAACGCCCGTGAGCGATGAAGGCCCTCGGGTCGTAAGCTCTGTCTCAAGGAAGATAATGACGGT
a498e9b5a610147b50de9176cee83849	410	AGTGGGGAATCTTGGCCAATGGGGGAAACCCCTGACGCGAGCAGCCCGTGCGGGACGAAGGCCCTCGGGTCGTAAGCTCTTTTATACGGGAATAAGTGTAGGC
54d8ed8275f89df38f123e98310d1ac3	413	AGTGGGGAATATTGCACAAATGGACCAAAAAGTCTGATCCAGCAATCTGTGTGCACGATGACGGTCTTAGGATTGTAAGTCTTTCAATTGGGGAAGAAAAAATGA
e18b13c0a48737c7e739e1f881c2c4b2	430	AGTGAGGAATATTGGTCAATGGGCGAGAGCCTGAACCAGCCAAAGTAGCGTGAAGGATGACTGCCCTATGGGTTGTAAGCTCTTTTATAAGGAATAAGTGTCCGGT
ba42c01bc3eb875307a7339eb249d9ca	413	AGTGGGGAATATTGCACAAATGGACCAAAAAGTCTGATCCAGCAATCTGTGTGCACGATGAAGTTTTTCGGGAATGTAAGTCTTTTATAGGGAATAAGTGTAGGC
5397cf19dbe23d4b7ea55e91901f6899	435	AGTGGGGAATATTGCACAAATGGGCGCAAGCCTGATGCAGCCATGCCCGTGATGAAGAAGGCCCTCGGGTTGTAAGTCTTTTATAGGGAATAAGTGTAGGC
4a32ec31f42357144d7cf0ab6956aae5	430	AGTGAGGAATATTGGTCAATGGGCGAGAGGCTGAACCAGCCAAAGTAGCGTGAAGGAAAGGATCTATGGGTTGTAAGCTCTTTTATAGGGAATAAGTGTAGGC
848eca9cd5555c4edf12439ecf66475f	410	AGTGGGGAATCTTGGCCAATGGGGGCAACCCCTGACGCGAGCAGCCCGTGCGGGACGAAGGCCCTCGGGTCGTAAGCTCTTTTATAGGGAATAAGTGTAGGC

**Abb. 6** Auszug aus der ASV Feature-ID Tabelle mit den zugehörigen Sequenzen. Diese sind mit einem Link unterlegt der zu NCBI nt-Datenbank (<https://www.ncbi.nlm.nih.gov/nucleotide/>) führt.

## 11. Exportieren von repräsentativen Sequenzen (ASVs)

Anlegen eines neuen Ordners

```
mkdir 03-asv-seqs
```

Exportieren von repräsentativen Sequenzen (ASVs)

```
qiime tools export
--input-path 02-qiime/rep-seqs.qza
--output-path 03-asv-seqs
```

Umbenennung von 03-asv-seqs/dna-sequences.fasta in 03-asv-seqs/ASV-sequences.fasta

```
mv 03-asv-seqs/dna-sequences.fasta 03-asv-seqs/ASV-sequences.fasta
```

## 12. Erstellen eines phylogenetischen Stammbaums

Bei der Erstellung eines phylogenetischen Stammbaums wurde ein Alignment der Sequenzen über das Plugin *MAFFT* (multiple sequence alignment program, <https://mafft.cbrc.jp/alignment/server/>) gemacht.

```
qiime alignment mafft
--i-sequences 02-qiime/rep-seqs.qza
--o-alignment 04-phylogenetic-tree/aligned-rep-seqs.qza
--p-n-threads 22
```

Danach wurde das Alignment gefiltert, um Lücken in den Sequenzen zu entfernen.

```
qiime alignment mask
--i-alignment 04-phylogenetic-tree/aligned-rep-seqs.qza
--o-masked-alignment 04-phylogenetic-tree/masked-aligned-rep-seqs.qza
```

Das Erstellen eines Stammbaumes erfolgte mit dem Plugin Fasttree (<http://www.microbesonline.org/fasttree/>). Der Stammbaum wurde dann entlang den beiden längsten Zweigen ausgerichtet (midrooting).

```
qiime phylogeny fasttree
--i-alignment 04-phylogenetic-tree/masked-aligned-rep-seqs.qza
--o-tree 04-phylogenetic-tree/unrooted-tree.qza
--p-n-threads 22
```

```
qiime phylogeny midpoint-root
--i-tree 04-phylogenetic-tree/unrooted-tree.qza
--o-rooted-tree 04-phylogenetic-tree/rooted-tree.qza
```

### 13. Alpha rarefaction Visualisierung

Um festzustellen, ob die Proben ausreichend tief sequenziert wurden, wurde eine Alpha-Rarefaction Kurve für alle Proben erstellt. Das Plugin „*diversity alpha-rarefaction*“ generiert hierfür eine Visualisierungsdatei.

Es wurde ein neuer Ordner erstellt, in dem die Alpha-Rarefraction Diagramme berechnet und abgespeichert werden.

```
mkdir 05-alpha-metrics/
qiime diversity alpha-rarefaction
--i-table 02-qiime/table.qza
--i-phylogeny 04-phylogenetic-tree/rooted-tree.qza
--p-max-depth 130000
--m-metadata-file sample-metadata.tsv
--o-visualization 05-alpha-metrics/alpha-rarefaction.qzv
```

### 14. Alpha-Diversität

Die Alpha-Diversität ist ein Maß für die mikrobielle Artenvielfalt und beschreibt die Anzahl der in einer Probe vorkommende Arten. Die Abundanzen der einzelnen Arten spielen dabei keine Rolle, so können zwei Proben dieselbe Alpha-Diversität aufweisen, aber eine unterschiedliche Mikrobiom Struktur haben (Estaki et al. 2020). Die Eingabe *--i-table* lädt die ASV-Tabelle. *--p-metrik* definiert die Alpha-Diversitätsparameter, die berechnet werden sollen (default: Chao1, Shannon, Simpson, observed features und goods coverage). Der Output *--o-alpha-diversity* gibt die Ausgabedateien an.

```
qiime diversity alpha
```

```
--i-table 02-qiime/table.qza
--p-metric ${i}
--o-alpha-diversity 05-alpha-metrics/${i}-metric.qza
```

Es wurde danach eine tabellarische Ansicht der Metadaten erzeugt (**Abb. 7**)

```
qiime metadata tabulate
--m-input-file 05-alpha-metrics/${i}-metric.qza
--o-visualization 05-alpha-metrics/${i}-metric.qzv
```

Sample ID #q2:types	shannon numeric
S15L	5.9569274729
S26H	5.852530583
S13L	5.8456018231
S12L	5.8219846428
S32H	5.8195557103
S30H	5.7502331943
S25H	5.738889431
S18L	5.733719101
S24H	5.657511717
S29H	5.5613320637
S06L	5.5402686355
S33H	5.5149826099
S20L	5.4743430806
S17L	5.4261629405

**Abb. 7** Ausschnitt aus der Ansicht im Qiime Viewer, welcher hier exemplarisch Werte des Shannon Parameters für ausgewählte Proben darstellt

## 15. Taxonomische Klassifizierung bis auf Genus Ebene

Der Befehl `qiime feature-classifier-classify-sklearn` wurde verwendet, um ASVs zu klassifizieren. Für den Input wurden taxonomische Datenbanken genutzt. Hier wurde „*Silva 132 99% OTUs full-length sequences*“ verwendet, welche zum freien Download verfügbar ist unter <https://data.qiime2.org/2019.7/common/silva-132-99-nb-classifier.qza>.

Die Visualisierung (**Abb. 8**) zeigt für jedes ASV die taxonomischen Namen von der Domäne bis hin zur Art. Zusätzlich sind Informationen zum Konfidenzniveau enthalten. Dies gibt an, mit welcher Wahrscheinlichkeit das ASV der zugeordneten Taxonomie entspricht.

Erstellen eines neuen Ordners

```
mkdir 06-taxonomy/
```

Zuweisen der Taxonomie zu den repräsentativen Sequenzen

```
qiime feature-classifier classify-sklearn
```

```
--i-classifier /Daten2/Peffi/silva-132-99-nb-classifier.qza
--i-reads 02-qiime/rep-seqs.qza
--o-classification 06-taxonomy/taxonomy.qza
--p-n-jobs 22
```

Erzeugen einer tabellarischen Ansicht der Metadaten

```
qiime metadata tabulate
--m-input-file 06-taxonomy/taxonomy.qza
--o-visualization 06-taxonomy/taxonomy.qzv
```

Feature ID #q2:types	Taxon categorical	Confidence categorical
2792fe47ad357f14ede6d1a00ab3390d	D_0__Bacteria;D_1__Cyanobacteria;D_2__Oxyphotobacteria;D_3__Chloroplast	0.9999999999448743
0cd2378b764a16a2dac9128f9f72e63c	D_0__Bacteria;D_1__Cyanobacteria;D_2__Oxyphotobacteria;D_3__Chloroplast	0.9999999999370063
10505c7b0c64ab9114dac08ab11fd84e	D_0__Bacteria;D_1__Bacteroidetes;D_2__Ignavibacteria;D_3__OPB56	0.9999999998760385
f3bcd55fdd2c52938180683c0f2cdfa0	D_0__Bacteria;D_1__Cyanobacteria;D_2__Oxyphotobacteria;D_3__Chloroplast	0.999999999843305
4276534c2e84754727cc046cf47c43e1	D_0__Bacteria;D_1__Cyanobacteria;D_2__Oxyphotobacteria;D_3__Chloroplast	0.9999999998230742
ae0772121b7627cce8ebd782fcb7d27a	D_0__Bacteria	0.9999999997959761
4a7575cc9872c7cfe929012d38d84aea	D_0__Bacteria;D_1__Spirochaetes;D_2__Spirochaetia;D_3__Spirochaetales;D_4__Spirochaetaceae;D_5__Treponema 2	0.9999999995224294
9340417acc30330592a55e7d8885b5d3	D_0__Bacteria;D_1__Firmicutes;D_2__Erysipelotrichia;D_3__Erysipelotrichales;D_4__Erysipelotrichaceae;D_5__Erysipelatoclostridium;D_6__Clostridiales bacterium CHKCI006	0.9999999993636379
c420156206cab9ca7fde3723cc71e30	D_0__Bacteria;D_1__Spirochaetes;D_2__Spirochaetia;D_3__Spirochaetales;D_4__Spirochaetaceae;D_5__Treponema 2	0.9999999993059598
49d4b563c38bb9390c45073486def002	D_0__Bacteria;D_1__Firmicutes;D_2__Clostridia;D_3__Clostridiales;D_4__Family XI;D_5__Peptoniphilus	0.9999999991902007
a44a7929f12bb6f354f28beb164671cd	D_0__Bacteria;D_1__Firmicutes;D_2__Negativicutes;D_3__Selenomonadales;D_4__Veillonellaceae	0.9999999981143642
d1a259c04a4762e440c0fb7323b12f2e	D_0__Bacteria;D_1__Spirochaetes;D_2__Spirochaetia;D_3__Spirochaetales;D_4__Spirochaetaceae;D_5__Treponema 2	0.9999999975301703
562f8e5fe4e773855042ddd638cddc26	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Bacteroidales;D_4__Bacteroidaceae;D_5__Bacteroides	0.999999997481666
aa1d9bb3f1f1d0f7e45c0f35b6ba6c16	D_0__Bacteria;D_1__Firmicutes;D_2__Erysipelotrichia;D_3__Erysipelotrichales;D_4__Erysipelotrichaceae	0.9999999974659423
8e5d59acc6a49e4fdcab833cc67b442	D_0__Bacteria;D_1__Bacteroidetes;D_2__Bacteroidia;D_3__Bacteroidales;D_4__Bacteroidaceae;D_5__Bacteroides	0.999999996297887
d2d2db22e9d45924cc375a26fba813a	D_0__Bacteria;D_1__Spirochaetes;D_2__Spirochaetia;D_3__Spirochaetales;D_4__Spirochaetaceae;D_5__Sediminispirochaeta;D_6__Spirochaeta sp. canine oral taxon 314	0.9999999960037655
50edfa82ed1a59b374f6143f08751b7a	D_0__Bacteria;D_1__Proteobacteria;D_2__Alphaproteobacteria;D_3__Rickettsiales;D_4__Anaplasmataceae;D_5__Wolbachia	0.999999995914066

**Abb. 8** Ausschnitt aus der tabellarischen Ansicht der ASV-ID und der zugehörigen taxonomischen Klassifizierung

## 16. Erstellen von Taxa-Bar-Plots

Aus dem taxonomischen Profil jeder Probe lässt sich nun mit dem Befehl „*qiime taxa barplot*“ ein interaktives Balkendiagramm erstellen. Die Balken können auf die gewünschten taxonomischen Ebene angepasst und nach Häufigkeit einer bestimmten taxonomischen Gruppe oder nach Metadaten-Gruppierungen sortiert werden (Hall und Beiko 2018).

Ordner Erstellen

```
mkdir 07-taxa-plots/
```

Mehrstufige Barplot-Visualisierung

```
qiime taxa barplot
```

```
--i-table 02-qiime/table.qza
--i-taxonomy 06-taxonomy/taxonomy.qza
--m-metadata-file sample-metadata.tsv
--o-visualization 07-taxa-plots/taxa-bar-plots.qzv
```

### 17. Beta Diversität

Die Beta Diversität gibt die Unterschiede zwischen den Proben an. Das Kommando „*Qiime Diversity core-metrics-phylogenetic*“ generiert neben der Alpha Diversität auch Beta-Diversitäts Distanzmatrizen wie z.B. die unweighted UniFrac-Distance. Als Input benötigt dieser Befehl die ASV Tabelle und den phylogenetischen Stammbaum. Für jede der Beta-Diversitäts-Metrik generiert QIIME2 automatisch eine Visualisierung. Die Visualisierungsdatei kann mit dem Tool *Emperor*, in dem Distanzmaßen angezeigt werden können, ausgewertet werden (Hall und Beiko 2018).

```
qiime diversity core-metrics-phylogenetic
--i-phylogeny 04-phylogenetic-tree/rooted-tree.qza
--i-table 02-qiime/table.qza
--p-sampling-depth 130000
--m-metadata-file sample-metadata.tsv
--output-dir 08-beta-diversity-depth_130000
```

### 18. Signifikanzberechnung der Alpha-Diversität

Die Proben werden auf signifikante Unterschiede in den Diversitätsparametern getestet. Die Eingabe „*qiime diversity alpha-group-significance*“ erstellt Boxplots mit der Diversität einer Metadatengruppe, die signifikanten Unterschiede der Gruppen werden mit dem *Kruskal-Wallis-Test* berechnet. *QIIME2* führt die statistischen Tests für jede in der Metadatei enthaltene Gruppe (in diesem Fall für den Vergleich zwischen Lymphoma/Gesund) durch.

Prüfung auf einen signifikanten Unterschied im Diversitätsparameter Shannon

```
qiime diversity alpha-group-significance
--i-alpha-diversity 08-beta-diversity-depth_130000/shannon_vector.qza
--m-metadata-file sample-metadata.tsv
--o-visualization 08-beta-diversity-depth_130000/shannon-vector-group-significance.qzv
```

Prüfung auf einen signifikanten Unterschied im Diversitätsparameter Evenness

```
qiime diversity alpha-group-significance
--i-alpha-diversity 08-beta-diversity-depth_130000/evennesse_vector.qza
--m-metadata-file sample-metadata.tsv
--o-visualization 08-beta-diversity-depth_130000/evennesse-vector-group-significance.qzv
```

### 19. Signifikanzberechnung der Beta Diversität

Der Test für die Beta Diversität kann mit einer PERMANOVA (Standardmethode) durchgeführt werden (Hall und Beiko 2018). Die weighted UniFrac-Distance Berechnung ist hierfür die Grundlage, sie ist ein quantitatives Maß für die Verschiedenheit der Diversität in einer Gemeinschaft, welches phylogenetische Beziehungen der ASVs miteinbezieht (Lozupone et al. 2007).

Berechnung der Weighted-UniFrac-Distance und der PERMANOVA

```
qiime diversity beta-group-significance
--i-distance-matrix 08-beta-diversity-depth_130000/weighted_unifrac_distance_matrix.qza
--m-metadata-file sample-metadata.tsv
--m-metadata-column Treatment
--o-visualization 08-beta-diversity-depth_130000/weighted-unifrac-subject-significance
```

### 20. Erstellen einer Tabelle mit den relative Abundanzen

Ein neuer Ordner für die relative Häufigkeiten der mikrobiellen Gemeinschaft wurde erstellt.

```
mkdir 09-relative-abundances
```

Die folgende Erklärung folgt der Durchführung zum Erstellen einer Tabelle auf Phylum Ebene (Level 2). Diese Schritte wurden jeweils von Level 2 bis auf Genus Ebene (Level 6) durchgeführt.

Es wurden ASV Feature-Tabelle mit der Taxonomie anstelle von Feature-IDs erstellt, wobei *table.qza* aus dem Output von DADA2 stammt und die Datei *taxonomy.qza* aus dem Schritt der taxonomischen Klassifizierung.

```
qiime taxa collapse
--i-table 02-qiime/table.qza
--i-taxonomy 06-taxonomy/taxonomy.qza
--p-level 2
--o-collapsed-table 09-relative-abundances/table-level-2.qza
```

Konvertierung der neue Abundanztabelle

```
qiime feature-table relative-frequency
--i-table 09-relative-abundances/table-level-2.qza
--o-relative-frequency-table 09-relative-abundances/table-level-2-relfreqs.qza
```

Exportierung der Daten in *biom*-Format

```
qiime tools export
--input-path 09-relative-abundances/table-level-2-relfreqs.qza
--output-path 09-relative-abundances/exported-table-level-2-relfreqs
```

Konvertierung der relativen Häufigkeitstabelle in eine Textdatei

```
biom convert
-i 09-relative-abundances/exported-table-level-2-relfreqs/feature-table.biom
-o 09-relative-abundances/exported-table-level-2-relfreqs/2-feature-table.txt --to-tsv
```

## 21. Konvertierung in eine Textdatei und hinzufügen der Taxonomie zur ASV Tabelle

Konvertierung von dem *biom*-Format in eine Textdatei

```
biom convert
-i 02-qiime/table_1/7ebbcad7-5bb9-49e6-ab98-7708aa64d474/data/feature-table.biom
-o 02-qiime/table_1/7ebbcad7-5bb9-49e6-ab98-7708aa64d474/data/feature-table.txt --to-tsv
```

Konvertierung der *table.qza* aus DADA2 in eine Tabelle mit relativen Abundanzen

```
qiime feature-table relative-frequency
--i-table 02-qiime/table.qza
--o-relative-frequency-table 09-relative-abundances/table-asv.qza
```

Konvertierung von *biom*-Format in eine Textdatei

```
biom convert
-i'/Daten2/Peffi/Cat_Lymphoma/QIIME2_conda/09-relative-abundances/table-asv/2657b1fc-41e7-4825-ab8e-3e638f7c5f9a/data/feature-table.biom'
-o'/Daten2/Peffi/Cat_Lymphoma/QIIME2_conda/09-relative-abundances/table-asv/2657b1fc-41e7-4825-ab8e-3e638f7c5f9a/data/asv-feature-table-relab.txt' --to-tsv
```

Hinzufügen der Taxonomie zur ASV Tabelle

```
cut-f109-relative-abundances/table-asv/2657b1fc-41e7-4825-ab8e-3e638f7c5f9a/data/asv-
feature-table-relab.txt > tmp1
tail -n+3 tmp1 > tmp2
for i in $(<tmp2); do grep "$i" 06-taxonomy/taxonomy/ca9764cf-2f56-4992-9187-
b326703cbd20/data/taxonomy.tsv >> tmp3; done
cut-f2 tmp3 > tmp4
tail -n+2 09-relative-abundances/table-asv/2657b1fc-41e7-4825-ab8e-3e638f7c5f9a/data/asv-
feature-table-relab.txt > tmp5
sed "1iTaxonomy" tmp4 > tmp6
paste tmp5 tmp6 > tmp7
mv tmp7 asv-relfreq-taxonomy.txt
```

### 3. Resultate

#### 3.1. Vermessung der DNA

Um die Qualität und Quantität der Proben zu verifizieren, wurden die DNA-Konzentrationen der Proben bestimmt (Qubit konz. ng/µl) (**Tab. 3**).

Proben-ID	Qubit konz. ng/µl	Gruppe
S1	71,6	Lymphoma
S2	52,4	Lymphoma
S3	5,08	Lymphoma
S4	14,8	Lymphoma
S5	16,1	Lymphoma
S6	76	Lymphoma
S7	7,72	Lymphoma
S8	94	Lymphoma
S9	36,3	Lymphoma
S10	19,6	Lymphoma
S11	19,9	Lymphoma
S12	16,6	Lymphoma
S13	12	Lymphoma
S14	13,6	Lymphoma
S15	0,5	Lymphoma
S16	3,6	Lymphoma
S17	20,8	Lymphoma
S18	9,28	Lymphoma
S19	44	Lymphoma
S20	38	Lymphoma
S21	78	Lymphoma
S22	6,56	Lymphoma
S23	72,8	Lymphoma
S24	20,6	Gesund
S25	15	Gesund
S26	5,72	Gesund
S27	11	Gesund
S28	2,04	Gesund
S29	23,8	Gesund
S30	17,5	Gesund
S31	3,49	Gesund
S32	0,149	Gesund
S33	10	Gesund
S34-LK	too low	Leerkontrolle

*Tab. 3 Probenübersicht über die Proben-ID, die gemessene DNA-Konzentration in ng/µl und die Zuordnung in die Lymphoma- oder gesunde Gruppe.*

### 3.2. Sequenzstatistik

Um zu überprüfen, ob die Proben ausreichend tief sequenziert wurden, wurden die Rarefaction Kurven im QIIME2 Viewer analysiert. Zu Beginn steigen die Kurven sehr stark an, bis sie anschließend ein Plateau erreichen. Jede der Proben erreicht mit zunehmender Sequenztiefe ein Plateau, was darauf hinweist, dass die Gesamtheit der Taxa erfasst wurde (Hall und Beiko 2018).

In **Tab. 4** ist die Sequenzanalyse der Proben zusammengefasst. Zu Beginn der Analyse lagen 11.006.277 Rohsequenzen vor. Durch die Qualitätsfilterung wurden 3.884.004 Sequenzen herausgefiltert, was 34,69 % der gesamten Rohsequenzen entspricht. Es blieben insgesamt 7.122.273 Sequenzen mit guter Qualität übrig (65,31 %), die für die weiteren Analysen verwendet werden konnten. Pro Probe befanden sich im Durchschnitt 323.714 Sequenzen (**Supplement Tab. 2, Supplement Abb.1**). Insgesamt wurden 3.005 ASVs detektiert. Im Durchschnitt waren die Sequenzen nach der Qualitätskontrolle und dem Merging der paired-end reads 413,5 bp lang.

	Rohsequenzen	Gefilterte Sequenzen	Gefilterte Sequenzen (%)	Kontrollierte Sequenzen guter Qualität	Kontrollierte Sequenzen guter Qualität (%)
<b>Anzahl</b>	11.006.277	3.884.004	34,69%	7.122.273	65,31%

*Tab. 4 Zusammenfassung der Sequenzanzahl der Rohdaten und der qualitätskontrollierten Daten.*

### 3.3. Die Alpha-Diversität

In **Tab. 5** sind die verschiedenen Alpha-Diversitätsparameter zusammengefasst. Von den Daten wurde jeweils der Mittelwert und die Standardabweichung berechnet und eine Unterteilung in die Gruppen gesund und Lymphoma vorgenommen.

Die observed ASVs sind ein quantitatives Maß für die Diversität einer mikrobiellen Gemeinschaft. Der Shannon Index schätzt den Reichtum und die Vielfalt anhand eines natürlichen Logarithmus. Dabei berücksichtigt er sowohl die Häufigkeit als auch die Gleichmäßigkeit der vorhandenen Taxa (Nagendra 2002). Der Simpson Index beschreibt die Spezies Diversität in einer Gemeinschaft. Er wird häufig genutzt, um die biologische Vielfalt eines Habitats zu beschreiben (Simpson 1949). Goods Coverage gibt die Deckung der Probe mit den bekannten Features wieder (1 bedeutet, dass 100% der Probe aus Features stammen, die mehr als einmal vorkommen) <https://mothur.org/wiki/coverage/> (Zugriff 18.05.21). Der Chao1 Index schätzt die Diversität einer Probe anhand von abundanten Daten und basierend

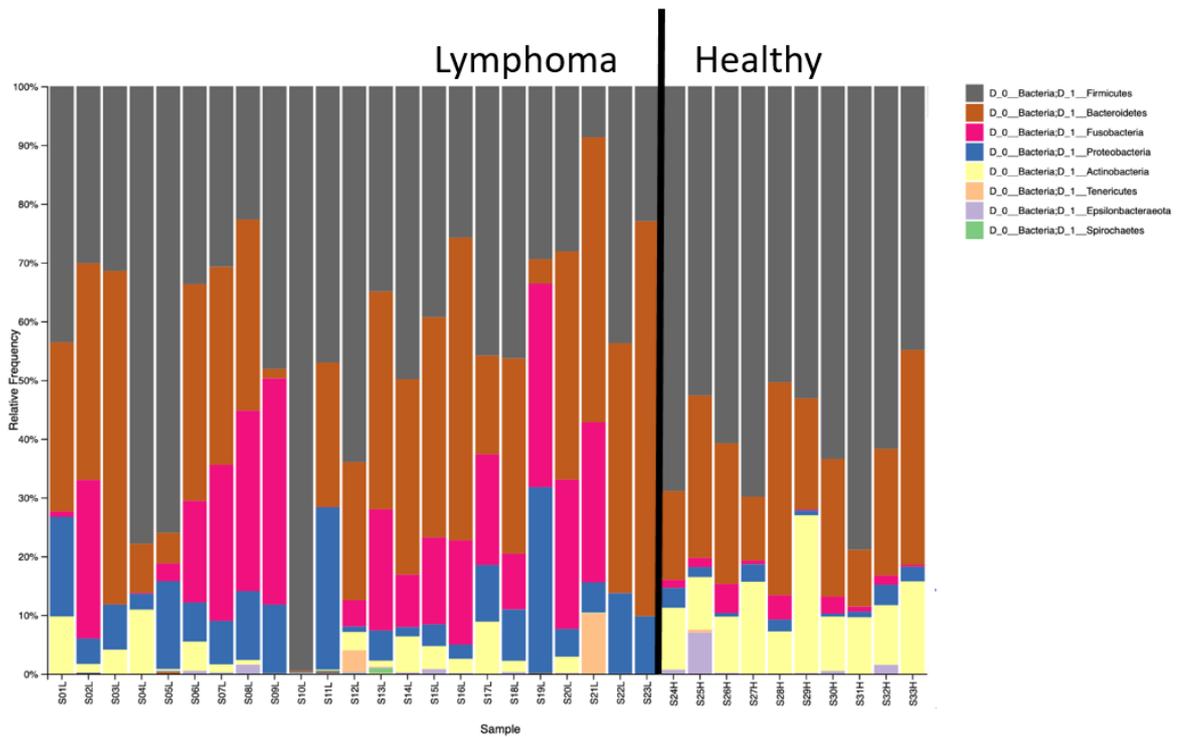
auf seltenen Features (Chao 1984). Aus **Tab. 5** geht hervor, dass die gesunden Tiere einen höheren Mittelwert in der Alpha-Diversität aufweisen als die Lymphoma Tiere. Die Standardabweichung ist bei den gesunden Tieren geringer als in der Lymphoma-Gruppe.

Gruppe	Mittelwert		Standardabweichung	
	Gesund	Lymphoma	Gesund	Lymphoma
<b>Observed Features</b>	279,2	209,26	59,92	101,34
<b>Shannon Index</b>	5,53	4,53	0,32	1,11
<b>Simpson Index</b>	0,95	0,88	0,02	0,10
<b>Goods Coverage</b>	1,00	1,00	<0,01	<0,01
<b>Chao 1 Index</b>	279,23	209,36	59,92	101,40

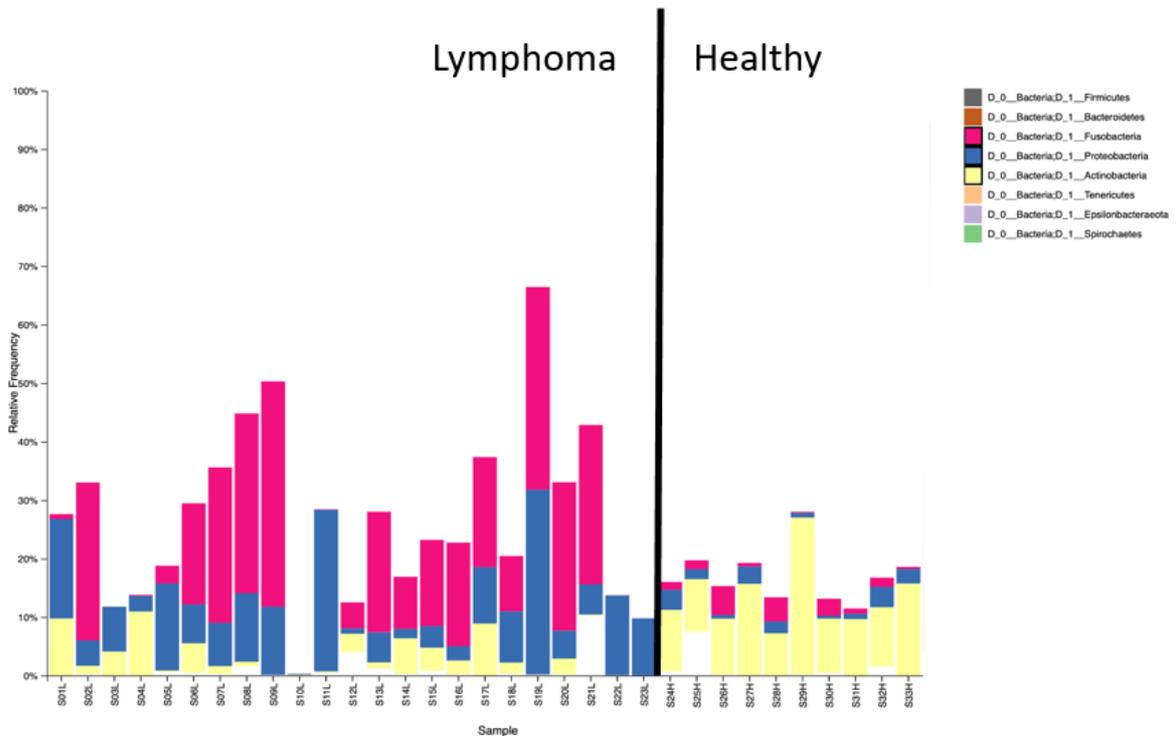
**Tab. 5** Überblick über die Alpha-Diversitäts Parameter. Dargestellt wird jeweils der Mittelwert, unterteilt in Gesund und Lymphoma und die Standardabweichung mit derselben Unterteilung.

### 3.4. Auswertung der Taxa-Bar-Plots

Die **Abb. 9** zeigt die relative Häufigkeit auf Phylum Level, sortiert nach aufsteigender Probennummer. Insgesamt wurden acht dominante Phyla ermittelt. In **Abb. 10** sieht man, dass neben *Firmicutes* und *Bacteroidota* in der Lymphoma-Gruppe *Fusobacteria* und *Proteobacteria* dominieren und der Gesamtanteil an *Actinobacteria* im Vergleich zu der gesunden Gruppe deutlich geringer ist.



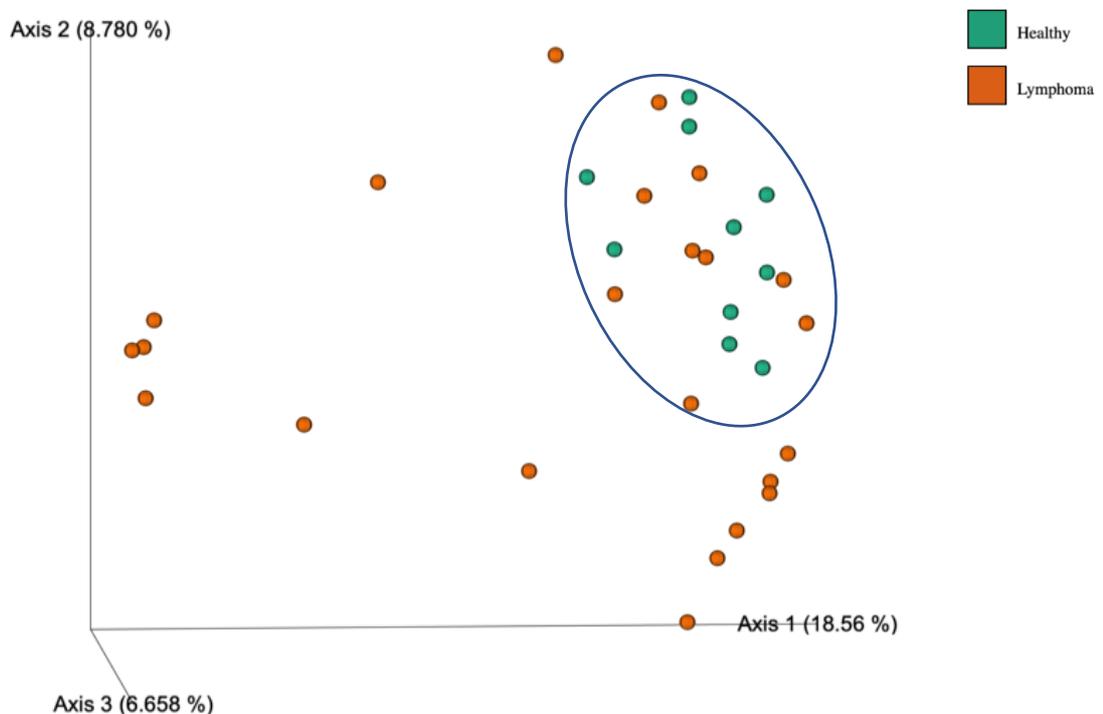
**Abb. 9** Taxa-Bar-Plot auf Phylum Ebene (Level 1) sortiert nach Probennummern S01L-S23L (L=Lymphoma) und S24H-S33H (H=Healthy/Gesund)



**Abb. 10** Modifizierter Taxa-Bar-Plot auf Phylum Ebene (Level 1), sortiert nach Probennummern S01L-S23L (L=Lymphoma) und S24H-S33H (H=Healthy/Gesund). Abgebildet sind zur besseren Übersicht nur die Phyla Fusobacteria, Proteobacteria und Actinobacteria.

### 3.5. Die Beta-Diversität

In **Abb. 11** sieht man in der unweighted UniFrac Analyse einen großen Unterschied der Proben, die zur Gruppe der erkrankten Katzen gehört (orange) im Gegensatz zu den gesunden Tieren (grün), welche eine geringere Distanz zueinander aufweisen. Neun erkrankte Tiere clustern mit den gesunden. Dieses Cluster ist in der Abbildung mit einer ovalen Umrundung gekennzeichnet (erkrankte Tiere im Cluster: S01L; S04L; S06L; S07L; S13L; S16L; S17L; S18L; S20L).



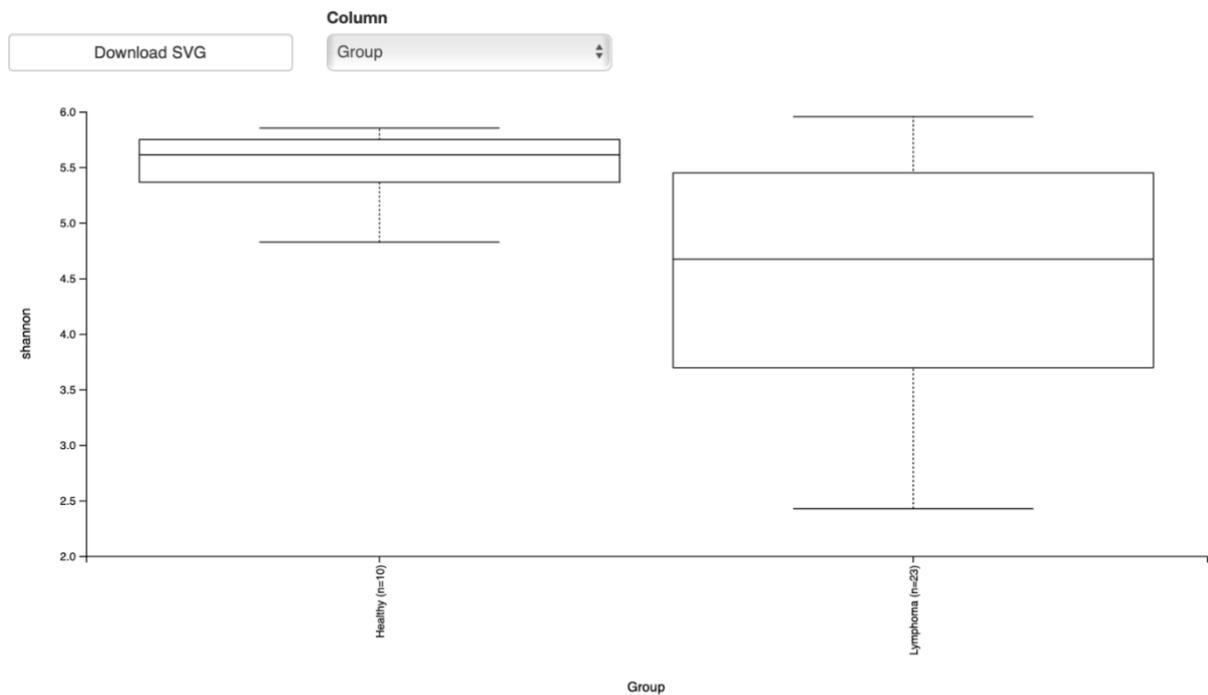
**Abb. 11** 3-D Emperor Abbildung der Proben, welche die unweighted UniFrac-Distance zeigt. Jeder Punkt stellt eine Probe dar

### 3.6. Die Signifikanzberechnung

#### 3.6.1. Alpha- und Beta-Diversität

In **Abb. 12** wird die gesunde Gruppe mit einem höheren Shannon-Wert assoziiert als die Lymphoma Gruppe. Es gibt einen signifikanten Unterschied der Diversität zwischen den beiden Gruppen ( $p$ -Wert = 0,015).

## Alpha Diversity Boxplots



[Download raw data as TSV](#)

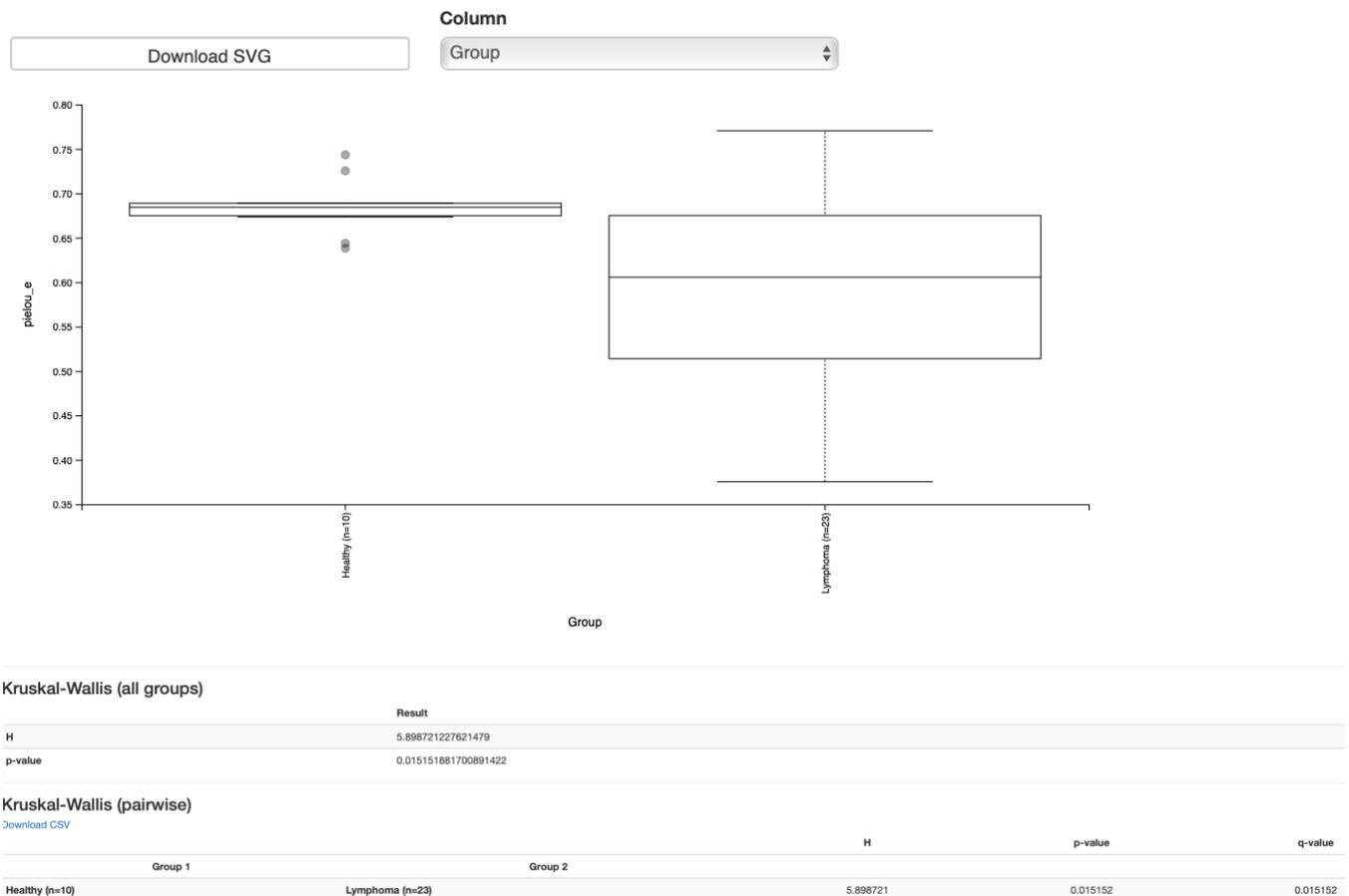
## Kruskal-Wallis (all groups)

	Result
H	5.898721227621479
p-value	0.015151881700891422

**Abb. 12** Screenshot der Ergebnisse der statistischen Analyse im Qiime Viewer. Gezeigt werden Boxplots für den Shannon Parameter in den Gruppen krank/gesund und die Ergebnisse des Kruskal-Wallis-Tests

Evenness ist ein Maß für die Gleichmäßigkeit der Gemeinschaft (Pielou 1966). Aus **Abb. 13** geht hervor, dass innerhalb der gesunden Gruppe eine geringere Diversität herrscht als in der Lymphoma-Gruppe. Dieser Unterschied ist signifikant (p-Wert=0,015).

## Alpha Diversity Boxplots



**Abb. 13** Screenshot der Ergebnisse der statistischen Analyse im Qiime Viewer. Gezeigt werden Boxplots für den Evenness Parameter in den Gruppen krank/gesund und die Ergebnisse des Kruskal-Wallis-Test

Auch in der PERMANOVA finden sich zwischen den Gruppen signifikante Unterschiede ( $p$ -Wert= 0,003).

### 3.7. Die relative Häufigkeit

Die **Abb. 14** zeigt eine Textdatei (.tsv), in der die relativen Häufigkeiten auf Phylum Ebene dargestellt sind, wobei die Summe jeder Probe/Spalte 100% ergibt. Aus dieser Tabelle lassen sich folgende Ergebnisse ablesen: Zum Beispiel, dass das fäkale Mikrobiom aus den Phyla *Firmicutes* (47,99 %), *Bacteroidota* (28,01 %), *Fusobacteria* (13,64 %), *Proteobacteria* (6,09 %) und *Actinobacteria* (2,51 %) besteht. Oder, dass die Probe S01L zu 43,57 % aus dem Phylum *Firmicutes*, zu 28,82 % aus *Bacteroidota* und zu 16,98 % aus *Proteobacteria* besteht und die Probe S02L sich aus 36,91 % *Bacteroidota*, 30,10 % *Firmicutes* und 26,97 % *Fusobacteria* zusammensetzt.

		Durchschnitt	S01L	S02L	S03L	S04L	S05L	S06L	S07L
D_0_Bacteria	D_1_Firmicutes	47,99%	43,57%	30,10%	31,41%	77,85%	75,98%	33,66%	30,70%
D_0_Bacteria	D_1_Bacteroidata	28,01%	28,82%	36,91%	56,81%	8,36%	5,24%	36,90%	33,71%
D_0_Bacteria	D_1_Fusobacteria	13,64%	0,86%	26,97%	0,02%	0,17%	3,03%	17,29%	26,58%
D_0_Bacteria	D_1_Proteobacteria	6,90%	16,98%	4,34%	7,66%	2,67%	14,92%	6,65%	7,40%
D_0_Bacteria	D_1_Actinobacteria	2,51%	9,75%	1,44%	4,11%	10,92%	0,20%	4,99%	1,34%
D_0_Bacteria	D_1_Tenericutes	0,73%	0,01%	0,00%	0,00%	0,00%	0,02%	0,00%	0,00%
D_0_Bacteria	D_1_Epsilonbacteraeota	0,14%	0,00%	0,07%	0,00%	0,01%	0,17%	0,51%	0,21%
Unassigned	—	0,03%	0,00%	0,16%	0,00%	0,00%	0,13%	0,00%	0,00%
D_0_Bacteria	D_1_Spirochaetes	0,05%	0,00%	0,00%	0,00%	0,00%	0,07%	0,00%	0,05%
D_0_Bacteria	D_1_Cyanobacteria	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
D_0_Bacteria	D_1_Patescibacteria	0,01%	0,00%	0,00%	0,00%	0,00%	0,24%	0,00%	0,00%
D_0_Bacteria	—	0,00%	0,01%	0,01%	0,00%	0,01%	0,00%	0,00%	0,00%
D_0_Bacteria	D_1_Synergistetes	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
D_0_Eukaryota	—	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
D_0_Bacteria	D_1_Chlamydiae	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
D_0_Archaea	—	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%

**Abb. 14** Ausschnitt aus der Tabelle, die die relative Häufigkeit dargestellt auf Phylum Ebene (Level 2) im Durchschnitt und von Probe S01L bis Probe S07L zeigt.

### 3.8. Die häufigsten ASVs

In **Tab. 6** sind die 30 häufigsten ASVs der Lymphoma-Gruppe und der gesunden Gruppe gelistet. Insgesamt beinhalten 59 % der Lymphoma Proben und 58 % der gesunden Proben die 30 häufigsten ASVs, wobei die Verteilung der ASVs sehr unterschiedlich ist.

ASV-ID	Durchschnitt Lymphoma	ASV-ID	Durchschnitt Gesund
2	8,46%	1	15,08%
1	6,50%	9	3,79%
3	4,82%	15	3,13%
5	4,01%	20	2,45%
8	3,77%	5	2,23%
6	3,55%	21	2,20%
4	2,52%	29	2,12%
13	2,27%	27	1,94%
12	2,27%	14	1,82%
11	1,79%	4	1,56%
17	1,71%	19	1,52%
18	1,69%	2	1,42%
7	1,55%	11	1,40%
16	1,55%	42	1,32%
28	1,12%	47	1,28%
14	1,11%	40	1,21%
10	1,04%	36	1,14%
9	1,03%	51	1,11%

34	1,02%	7	1,07%
26	0,98%	54	1,07%
24	0,95%	48	1,01%
30	0,92%	37	0,96%
31	0,89%	39	0,90%
22	0,80%	53	0,86%
25	0,71%	38	0,86%
46	0,66%	55	0,85%
41	0,64%	56	0,84%
19	0,61%	49	0,80%
52	0,60%	81	0,79%
63	0,58%	80	0,74%
<b>Summe</b>	<b>60,12 %</b>		<b>57,46 %</b>

*Tab. 6 Die 30 häufigsten ASVs der Lymphoma- und der gesunden Gruppe. Die dazugehörige Taxonomie kann aus Supplement Tab. 5 abgelesen werden.*

## 4. Diskussion

Ziel dieser Arbeit war es, einen NGS Datensatz mit dem bioinformatischen Analysetool QIIME2 auszuwerten und die Daten zu interpretieren. Dabei sollte folgende Hypothese untersucht werden: Die an Lymphomen erkrankte Katzen zeigen einen nachweisbaren Mikrobiota Shift (Dysbiose) im Kot und unterscheiden sich im Mikrobiota-Profil signifikant von gesunden Katzen.

### 4.1. Allgemeines zur Sequenzierstrategie

In den letzten drei Jahrzehnten gab es eine enorme Wissenssteigerung im Bereich der Genomik, ausgehend vom Human Genom Projekt, das 1990 gestartet wurde. Ziel des Projekts war es, das komplette menschliche Genom zu sequenzieren. Zur Umsetzung wurden knapp drei Milliarden US-Dollar über einen Zeitraum von 13 Jahre investiert („Human Genome Project FAQ“). Heutzutage liegen die Kosten für die Sequenzierung eines vollständigen Bakteriengenoms unter 1.000 US-Dollar <https://www.nanoporetech.com> (Zugriff 20.05.2021). Auch was die Dauer des Sequenziervorgangs betrifft, gab es deutliche Fortschritte. Die Länge der Sequenzen, die mit NGS Methoden generiert wird, ist unterschiedlich. So werden beispielsweise bei einer MiSeq Illumina Sequenzen maximal 300 bp Sequenzen generiert. Umso länger die Sequenz, desto eindeutiger ist die taxonomische Klassifizierung (Venter 2010). Ein Problem, das z.B. bei der Sequenzierung durch Illumina entstehen kann, ist, dass die Synthese aufgrund fehlerhaften Clusters nicht synchron abläuft und es dadurch zu einer Verringerung von Zyklen kommen kann. Des Weiteren muss darauf geachtet werden, die Flow-Cell nicht mit Proben-DNA zu überladen, da dies sonst zu einem „Überclustern“ führen kann. Beides führt zu einem Qualitätsverlust und zu einer erhöhten Fehlerrate der Probe (Chen et al. 2017). Für diese Arbeit wurde MiSeq von Illumina ausgewählt, da dieses Instrument mit der Erstellung von paired-end Sequenzen (2x300bp) vergleichsweise lange Sequenzlänge innerhalb der NGS Geräte der zweiten Generation aufweist und eine hohe Sequenzanzahl erzielt wird, um die hohe Diversität von Kotproben abzudecken (**Tab. 1**). Wie bereits erwähnt, stellt die Nutzung von NGS und speziell die der Illumina-Technologie eine enorme Zeitersparnis bei hoher Sensitivität und geringen Kosten pro Base dar. MiSeq ist derzeit die Methode der Wahl zur Erfassung der mikrobiellen Diversität.

Die dritte Generation der NGS, zu der PacBio (Pacific Biosciences) Sequenzierungen gehören, bietet deutlich längere Leselängen mit noch schnellerer Durchlaufgeschwindigkeit. Die Nachteile dieser Methode sind ein geringer Durchsatz, eine höhere Fehlerrate und höhere Kosten pro Base. Da sich die Verfahren der zweiten und dritten NGS Generation ergänzen,

wird immer häufiger eine Kombination (Hybrid-Strategie) aus beiden Methoden genutzt, um die Genauigkeit der Datenanalyse zu erhöhen (Thürmer 2014, Rhoads und Au 2015, Slatko et al. 2018).

## 4.2. Bioinformatische Analysetools

Derzeit gibt es verschiedene leistungsstarke bioinformatische Analysetools, die auf unterschiedlichen Methoden basieren. Zum einen die Tools, welche auf einer Kommandozeile beruhen, wie z.B. Mothur (Schloss et al. 2009) und DADA2 (Callahan et al. 2016). Dabei werden Befehle in Text-Form in das Programm eingegeben und ausgeführt. Des Weiteren gibt es QIIME und QIIME2 (Caporaso et al. 2010, Bolyen et al. 2019b), welche Python-basiert sind. Python ist eine universelle Programmiersprache, für welche bei der Verwendung von QIIME und QIIME2 ein gewisses Verständnis erfordert wird. Außerdem gibt es noch webbasierte Tools wie Calypso (Zakrzewski et al. 2017). Generell entwickeln sich die Analysetools im Bereich der Mikrobiomforschung sehr schnell weiter und die Qualität der Software kann stark variieren. Daher ist es wichtig, geeignete Programme zu finden oder zu entwickeln, die einen gewissen Qualitätsstandard aufweisen. Ziel ist es, dass weder Software noch Analysetools die Ergebnisse der mikrobiellen Zusammensetzung der Proben beeinflussen. Zusätzlich ist es wichtig, dass die Tools benutzerfreundlich und unabhängig von dem bioinformatischen Vorwissen sind, sowie reproduzierbare Ergebnisse generieren und einen geringen Zeitaufwand für die Analyse in Anspruch nehmen.

## 4.3. Vor- und Nachteile der QIIME2 Analyse

QIIME2 ist eine vielfältige Open-Source-Plattform, die auf der Grundlage einer Plugin-Architektur entwickelt wurde und schnell an individuelle Anforderungen angepasst werden kann. Zum Beispiel bietet QIIME2 eine Qualitätskontrolle der Daten über das Plugins DADA2 (q2-dada2) an. DADA2 ist ein Open-Source-Softwarepaket zur Korrektur von Amplikonfehlern in Illumina-Sequenzen. Es fügt auch Sequenzen mit gepaarten Enden (paired-end) zusammen und derepliziert Sequenzen. Dabei werden identische Sequenzen zusammengefasst und mit Einbezug der Häufigkeit als eindeutige/repräsentative Sequenzen gekennzeichnet. Diese Art der Zusammenfassung reduziert die anschließende Rechenleistung erheblich (Callahan et al. 2016). Ein Beispiel für die individuelle Anpassung des Tools ist die Eingabe „DADA 2 denoising-paired“. Dazu werden zwei erforderliche Größen benötigt. Zum einen --p-trunc-len-

f, welcher die Position angibt, an der die Vorwärtssequenz abgeschnitten wird und zum anderen --p-trunc-len-r, an dem die Rückwärtssequenz abgetrennt wird. Ein optionaler Parameter ist: --p-max-ee. Der Standard liegt bei 2 und gibt die maximale Anzahl an erwarteten Fehlern in einer Sequenz an, bevor diese verworfen wird. Um Chimären in den Daten zu verhindern, ist das Entfernen der Primer nötig. Dazu dienen folgende Eingaben: --p-trim-left-f und --p-trim-left-r, welche auf die Länge der Primersequenzen eingestellt werden und diese vor der Qualitätskontrolle entfernt (Hall und Beiko 2018). Falls Nutzer keine individuellen Anpassungen tätigen möchten, kann jederzeit auf Standardeinstellungen zurückgegriffen werden. QIIME2 bietet ein neues interaktives Visualisierungstool, den „*QIIME2 Viewer*“, an, welches ohne die vorherige Installation der Software genutzt werden kann. Fehlerhafte Ergebnisse können von QIIME2 erkannt und zurückverfolgt werden. Das Weiteren ist es möglich, zu einem späteren Zeitpunkt nachzuvollziehen, wie die Ergebnisse generiert wurden. Die Reproduzierbarkeit, die Transparenz und die Klarheit sind die Grundlagen beim QIIME2-Design und tragen damit zu einem großen Fortschritt bei der Datenverarbeitung bei (Bolyen et al. 2019b).

Heutzutage liefern verschiedene Sequenzierungstools noch variierende phylogenetische und taxonomische Verteilungen und beinhalten unterschiedliche Ansätze der Qualitätskontrolle (Allali et al. 2017). Häufig wird die bioinformatische Analyseplattform QIIME2 mit Mothur verglichen. Der wesentliche Unterschied zwischen den beiden Tools liegt darin, dass in QIIME2 ein Schwerpunkt auf die Visualisierung der Daten gelegt wird, wie z.B. mit dem *QIIME2 View*, während Mothur auf das Prozessieren von Daten spezialisiert ist und für die Veranschaulichung der Daten weiterführende Tools nötig sind (Bolyen et al. 2019b).

Eine Herausforderung für alle Tools ist die Bildung von Sequenzchimären während der Sequenzierung. Chimären werden während des PCR-Amplifikationsprozess gebildet und sind falsche Rekombinationen. Es entsteht der fälschliche Eindruck, dass diese Sequenzen von einem einzelnen mikrobiellen Organismus stammen. Darum ist das Erkennen und Entfernen von Chimären (über UCIME & VSEARCH) bedeutsam (Buza et al. 2019). QIIME2 bietet verschiedene Methoden, um die wahrscheinlichste taxonomische Zugehörigkeit der ASVs vorherzusagen. Der Befehl *qiime feature-classifier-classify-sklearn* wird verwendet, um Sequenzen zu klassifizieren (Hall und Beiko 2018). Dazu nutzt die Software die maschinell lernende Python-Bibliothek *scikit-learn* (Pedregosa et al. 2011). Wie bereits im Ergebnisteil der taxonomischen Klassifizierung erwähnt, gibt es QIIME2 Klassifikatoren, welche benutzerdefiniert „trainiert“ werden können (q2-Feature-Classifier-Protokoll <https://docs.qiime2.org/2021.2/tutorials/feature-classifier/> (Zugriff 27.05.21)) und somit eine genauere Klassifikation ermöglichen (Bokulich et al. 2018). Dabei werden Informationen über die spezifische Probenvorbereitung, die Länge der Sequenzen und die verwendeten Primerpaare hinterlegt. Die Standarddatenbanken, welche zum größten Teil öffentlich

verfügbar sind, wie beispielsweise SILVA oder Greengenes, sind allerdings genauso in QIIME2 anwendbar. Die Ergebnisse über die Datenbanken sind im Allgemeinen genauer als die Ergebnisse der einfachen BLAST-Suche, die mit der *rep-seqs.qzv*-Visualisierung (**Abb. 6**) verknüpft sind.

QIIME2 ist die Weiterentwicklung und Verbesserung von QIIME1 und stellt einen enormen Fortschritt in der Benutzerfreundlichkeit dar. Wissenschaftler schätzen zum einen, dass durch die klare Struktur von QIIME2 der Überblick während der Datenanalyse erleichtert wird und die Daten jederzeit zu ihrem Ursprung zurückverfolgt werden können. Zum anderen schätzen sie, dass durch die Pluginstruktur das Tool unkompliziert, vielfältig nutzbar und leicht an die persönlichen Anforderungen anpassbar ist. Besonders bei QIIME2 ist die direkte Zusammenarbeit von Entwickler und Nutzer positiv hervorzuheben. Bei Unklarheiten steht das QIIME2-Forum zur Verfügung, in welchem Fragen schnell, klar und meist innerhalb von 1-2 Werktagen beantwortet werden. Das Teilen von Daten (.qzv), zum Beispiel per E-Mail, ist deutlich einfacher dadurch geworden, dass keine vorherige Installation des *QIIME2 Viewer* nötig ist. Des Weiteren ist die Datenanalyse ohne bioinformatisches Vorwissen oder Programmierungserfahrung möglich. Für viele Autoren ist QIIME2 zur Zeit das beste frei zugängliche Tool und bietet transparente Mikrobiomforschung, was einen großen Fortschritt für die Reproduzierbarkeit von Daten darstellt (Bolyen et al. 2019b).

#### 4.4. Überblick über die Analyse des Lymphoma Datensatzes

Die Analyse des Lymphoma Datensatzes hat ergeben, dass es einen signifikanten Unterschied des fäkalen Mikrobioms zwischen den beiden Gruppen (Lymphoma/Gesund) gibt. Somit konnte die ursprüngliche Hypothese bestätigt werden.

Es ist interessant, dass die Proben der gesunden Gruppe eine hohe Diversität innerhalb der einzelnen Proben, aber einen geringen Unterschied zwischen den Proben der Gruppe aufweisen, obwohl die Gruppe „Gesund“ nicht standardisiert ist (unterschiedliches Futter, Umwelteinflüsse variieren, unterschiedlicher Kontakt zu anderen Tieren und Menschen, unterschiedliches Alter und Geschlecht). Die Lymphoma-Gruppe weist innerhalb der einzelnen Proben eine geringere Diversität auf, sie unterscheiden sich jedoch untereinander stark. **Abb. 11** kann entsprechend der Metadatenkategorie eingefärbt und die 3D-Abbildung gedreht werden, um festzustellen, ob die Proben zusammen clustern oder nicht. Dabei sieht man, dass neun Tiere der Lymphoma-Gruppe (S01L, S04L, S06L, S07L, S13L, S16L, S17L, S18L und S20L) mit den Tieren aus der gesunden Gruppe zusammenfallen. Dies könnte bedeuten, dass sich das Mikrobiom dieser Tiere noch nicht verändert hat oder gewisse mikrobielle Phylotypen dazu beitragen, dass das Mikrobiom trotz Erkrankung stabil bleibt. In **Supplement Tab. 3** sind

die Proben und ihre Metadaten aufgelistet. Was auffällt ist, dass keine der Lymphoma-Katzen, welche eine Behandlung erfahren hat (S03L, S14L und S15L), mit den gesunden Proben zusammenfällt. Nur die unbehandelten Tiere zeigen ein Clustern mit der Kontrollgruppe. Grund für diese Häufung könnte zum einen sein, dass die Krankheit (Lymphom) in einem früheren Stadium erkannt wurde und noch keine Auswirkung auf die Zusammensetzung des Mikrobioms hat. Zum anderen könnte eventuell die Lokalisation des Lymphoms keine oder zu geringe Änderungen des Mikrobioms verursachen, die über fäkale Proben nicht nachgewiesen werden können. Um diese Fragen beantworten zu können, müsste ein Datensatz mit einer größeren Probenanzahl analysiert werden. Es gibt erst einige wenige Publikationen darüber, wie sich das Mikrobiom von gesunden Katzen verhält und bis dato noch kaum Informationen darüber, wie es sich bei einer Lymphom Erkrankung verändert.

#### 4.4.1. Taxonomischer Vergleich mit der aktuellen Literatur

Frühere Untersuchungen zum Mikrobiom der Katze, bei denen traditionelle Kulturverfahren genutzt wurden, haben einen überwiegenden höheren Anteil an fakultativen und obligat anaeroben Bakterien im Gegensatz zu anderen Säugetieren gezeigt (Johnston et al. 1993, Papsoulotis et al. 1998). Studien, in denen mit Gewebeproben aus dem Darm von Katzen gearbeitet wurde, ergaben, dass 99 % des Darmmikrobioms aus den Phyla *Firmicutes*, *Bacteroidota*, *Proteobacteria*, *Fusobacteria*, und *Actinobacteria* bestehen (Suchodolski 2011b). Die Mehrheit der Mikrobiota wurde als *Firmicutes* klassifiziert, gefolgt von *Proteobacteria*, *Bacteroidota*, *Fusobacteria* und *Actinobacteria* (Ritchie et al. 2008).

In dieser Arbeit besteht das Mikrobiom der gesamten Proben zu 99 % aus folgenden Phyla: *Firmicutes* (47,99 %), *Bacteroidota* (28,01 %), *Fusobacteria* (13,64 %), *Proteobacteria* (6,90 %) und *Actinobacteria* (2,51 %) (**Supplement Tab. 1**). Diese Ergebnisse stimmt mit denen von Suchodolski et al. überein (Suchodolski 2011b). In der Studie von Meazzi et al. (Meazzi et al. 2019) war *Firmicutes* auch am häufigsten nachweisbar, gefolgt von *Bacteroidota*, *Actinobacteria* und *Proteobacteria*. *Firmicutes* war ebenfalls in der Studie von Desain et al. (Desai et al. 2009) am häufigsten vertreten, gefolgt von *Actinobacteria*, *Bacteroidota* und *Proteobacteria*. Auch Barry et al. (Barry et al. 2012) beschreiben in ihrer Studie *Firmicutes* (36–50 %) als dominantes Phylum, gefolgt von *Bacteroidota* (24–36 %), und *Proteobacteria* (11–12 %). Ebenso wie Deng und Swanson (Deng und Swanson 2015), wo auch *Firmicutes* dominierte, an zweiter Stelle *Bacteroidota*, gefolgt von *Fusobacteria*, *Actinobacteria* und *Proteobacteria*. In der Studie von Handl et al. (Handl et al. 2011) ist ebenfalls *Firmicutes* (92 %) am häufigsten vertreten, gefolgt von *Actinobacteria* (7,3 %), *Bacteroidota* (0,5 %) und *Fusobacteria* (0,04 %). Nur in der Studie von Tun et al. (Tun et al. 2012) dominierte *Bacteroidota* (68 %), gefolgt von *Firmicutes* (13 %), *Proteobacteria* (6 %), *Actinobacteria* (1 %)

und *Fusobacteria* (0,7 %). Dieser Unterschied könnte erklärt werden durch i.) die geringe Probenanzahl von n=5, wobei alle Katzen in dieser Studie Wohnungskatzen (Asien) waren, ii.) das Fehlen eines „beads beater“ Schrittes während der DNA-Extraktion, wo widerstandsfähige Bakterien wie *Firmicutes* nicht aufgebrochen werden konnten.

Die Dominanz von *Firmicutes* in vielen Studien sowie auch in dieser, zeigt, dass für die Katze die Verdauung von an sich unverdaulichen Ballaststoffen sehr effizient möglich ist und damit kurzkettige Fettsäuren zur Verfügung stehen. Einfache Zuckermoleküle werden durch das Überangebot an *Firmicutes* zur Verfügung gestellt und sichern so, dass das Maximum an verfügbarer Energie bereitsteht.

Aus der frühen Literatur wissen wir, dass die am häufigsten isolierten Bakteriengruppen aus dem GI-Trakt *Bacteroides* spp., *Clostridium* spp., *Enterococcus* spp., *Streptococcus* spp., *Fusobacteria* spp. und *Eubacteria* spp. sind (Minamoto et al. 2012a). Die Datenanalyse dieser Arbeit ergab dominante Familien zugehörend zu *Bacteroidaceae* (17,97 %), *Peptostreptococcaceae* (14,99 %), *Lachnospiraceae* (11,42 %), *Fusobacteriaceae* (10,44 %), *Ruminococcaceae* (6,60 %), *Coriobacteriaceae* (4,78 %), *Prevotellaceae* (4,77 %), *Clostridiaceae* 1 (4,64 %), *Burkholderiaceae* (3,61 %) und *Enterobacteriaceae* (2,41 %) (**Supplement Tab. 6**). Auf Genus- Ebene gibt es große Unterschiede zwischen schon publizierten Studien, was die extreme Variabilität und Anpassungsfähigkeit des Darmmikrobioms an unterschiedliche Umweltfaktoren widerspiegelt (Meazzi et al. 2019). Wie in der Einleitung bereits erwähnt, gibt es aber auch bei den Analysetools große Unterschiede in der taxonomischen Zuordnung ab der Familien-Ebene aufwärts. Dies führt zu unterschiedlichen Beta-Diversitätsberechnungen und erschwert dadurch die Vergleichbarkeit der Studien auf Familien-, Genus- und Spezies-Ebene. Speziell auf Genusebene (**Supplement Tab. 7**) ist auch in diesem Datensatz manchmal keine Zuordnung mehr möglich, was mit anderen schon publizierten Datensätzen, welche mit MiSeq sequenziert wurden, übereinstimmt (Zwirzitz et al. 2020).

Das Mikrobiom ist ein empfindliches System, das sich je nach Lebenssituation anpasst und verändert. Die Unterschiede zwischen den Ergebnissen der Studien könnten auf Unterschiede in der Ernährung, dem Alter, der Körperkondition, den Haltungsbedingungen, einer Kastration, dem Vorhandensein von entzündlichen Erkrankungen und vielen anderen Parameter zurückzuführen sein (Deusch et al. 2015). Vorangegangene Studien zeigen einen nicht zu vernachlässigenden Unterschied des Mikrobioms zwischen unterschiedlichen Altersgruppen (Suchodolski et al. 2015). Ziel ist es eine möglichst homogene Gruppe an Katzen für weiterführende Studien zu gewinnen, um die Aussagekraft der Ergebnisse zu erhöhen.

## 4.5. Das Darmmikrobiom der Katze im gesunden Zustand und während der Lymphomerkrankung

Im Folgenden Kapitel wird genauer auf die Bakterien, deren Aufgaben und Vorkommen eingegangen, sowie diese diskutiert. Jedes Tier zeigt ein individuelles mikrobielles Profil, dennoch gehört die überwiegende Mehrheit eines gesunden Darmmikrobioms zu den Phyla der *Firmicutes* und *Bacteroidota*. Erweitert werden diese von dem Phylum der *Actinobacteria*, welches in deutlich geringerer Anzahl vorkommt. Im Falle einer Krankheit kommt es häufig zu einer Dysbiose. Meist geht damit eine Verringerung der Bakterienvielfalt und Reduktion der obligat anaeroben Arten mit einer Zunahme der fakultativ anaeroben Bakterien einher (Walker und Lawley 2013). In dieser Arbeit besteht das gesamte Mikrobiom der gesamten Proben zu 99 % aus (**Supplement Tab. 5**): *Firmicutes* (47,99 %), *Bacteroidota* (28,01 %), *Fusobacteria* (10,45 %), *Proteobacteria* (6,90 %) und *Actinobacteria* (5,69 %). Vergleicht man die generierten Phyla der Lymphoma-Gruppe mit den der gesunden Gruppe, erkennt man, dass jeweils *Firmicutes* und *Bacteroidota* dominieren und in der gesunden Gruppe knapp 83 % des gesamten Mikrobioms ausmachen, während es in der Lymphoma-Gruppe nur 73 % sind. Dort stehen an dritter Stelle die *Fusobacteria*, an vierter Stelle die *Proteobacteria* und an fünfter Stelle die *Actinobacteria*. In der gesunden Gruppe stehen an dritter Stelle die *Actinobacteria* gefolgt von den *Proteobacteria* und knapp dahinter, an fünfter Stelle *Fusobacteria*.

### 4.5.1. *Firmicutes*-assoziierte Bakterien

*Firmicutes* ist das dominante Phylum des Darmmikrobioms der Katze und macht den größten prozentualen Anteil aus. Die grampositiven Bakterien sind obligat anaerob, dies bedeutet, dass sie nur in Abwesenheit von Sauerstoff überleben und sich vermehren können. Die Bakterien haben eine immunomodulatorische Wirkung und sind indirekt entzündungshemmend, indem sie kurzkettige Fettsäuren (SCFAs) produzieren. Die SCFAs setzen sich zum größten Teil aus Butyrat, aber auch Acetat und Propionat zusammen. Insbesondere die Dickdarmepithelzellen (Colonozyten) decken 60-70% ihres Energiebedarfs durch Butyrat (Walker und Lawley 2013). Der Zusammenhang von Butyrat produzierenden Bakterien und die damit einhergehende schützende Wirkung gegen Darmkrebs ist durch eine Vielzahl an Studien belegt (Orchel et al. 2005, Comalada et al. 2006, Tan et al. 2008).

Die Lymphoma Katzen weisen eine Reduktion von 17,87 % des Phylums *Firmicutes* auf. Auch andere Studien aus der Humanmedizin zeigen, dass es unabhängig von der genauen Art der Krankheit auch bei entzündlichen oder immunregulierten Darmerkrankungen oft zu einer Reduktion des Phylums *Firmicutes* kommt (Frank et al. 2007, Chang et al. 2008, Major und

Spiller 2014). Diese Dysbiose moduliert unter anderem auch die Fähigkeit zur Modifikation von Gallensäuren und wird aktuell als Krankheitsmarker diskutiert (Ogilvie und Jones 2012).

Die gesunde Gruppe weist einen 4 % höheren Anteil an *Ruminococcaceae* auf. *Ruminococcaceae* sind 7 $\alpha$ -dehydroxylierende Bakterien, welche die Reduktion von primären Gallensäuren zu sekundären Gallensäuren fördern. Die sekundären Gallensäuren weisen im Darm eine entzündungshemmende Wirkung auf. Somit entsteht durch eine Verringerung der *Ruminococcaceae*-Familie indirekt ein entzündungsförderndes Milieu. Ansätze zum Ausgleich der Dysbiose und therapeutischen Nutzung der *Ruminococcaceae*-Familie haben Potenzial in der Zukunft (Sinha et al. 2020).

In der Lymphoma-Gruppe weist die *Lachnospiraceae*-Familie einen Rückgang von 8 % im Gegensatz zu der gesunden Gruppe auf. Diese Reduktion der Mitglieder der *Lachnospiraceae* wird in vorangegangenen Studien mit entzündlichen Darmerkrankungen in Zusammenhang gebracht (Frank et al. 2007). Desweiteren konnte in einer Studie mit Mäusen, bei einem Darmkrebsmodell, die Polypenbildung an der Darmschleimhaut mithilfe von *Lachnospiraceae* verhindert werden (Mai et al. 2007). Dies lässt auf eine wesentliche Rolle der *Lachnospiraceae* bei der Aufrechterhaltung der Darmhomöostase schließen (Frank et al. 2007).

In vorangegangenen Studien konnte ein Anstieg der *Erysipelotrichaceae*-Familie bei entzündlich bedingten Magendarm Erkrankungen, wie zum Beispiel Darmkrebs, festgestellt werden (Chen et al. 2012, Zhu et al. 2014). Die Lymphoma-Daten dieser Arbeit ergeben keinen Anstieg der *Erysipelotrichaceae*, sondern eine Reduktion um 2 %. Einige Bakterienfamilien, zum Beispiel Mitglieder der *Erysipelotrichaceae*, werden mit einem Überzug an IgA-Immunglobulinen, unterschiedlicher Konzentrationen, in Zusammenhang gebracht (Kaakoush 2015). Durch IgA wird die Entzündung identifiziert und von den Bakterien anschließend gefördert. Dadurch wird der Schweregrad entzündlich bedingter Darmerkrankungen gesteigert (Palm et al. 2014). Diese Beobachtung kann durch die Lymphoma-Daten dieser Arbeit nicht bestätigt werden. Ein Grund dafür könnte eine vorangegangene Immunsuppression, speziell mit einer Verringerung der IgA Immunglobulinen, sein.

#### 4.5.2. *Bacteroidota* -assoziierte Bakterien

*Bacteroidota* sind gramnegative, obligat anaerobe Bakterien. Mitglieder des Phylums *Bacteroidota* sind metabolisch hoch flexibel und können sich schnell an veränderte Umweltbedingungen anpassen (Kolmeder et al. 2015). Im gesunden Darmmikrobiom sind Vertreter der *Bacteroidota* sehr häufig zu finden. Diese tragen zum Beispiel zur Bildung von Vitamin K<sub>2</sub> (Ramotar et al. 1984) und SCFs, vor allem Acetat und Propionat, bei (Fei und Zhao 2013). Durch eine ballaststoffreiche Ernährung können bis zu 10 % des täglichen Kalorienbedarfs des Wirts durch SCFAs gedeckt werden, welche durch die Fermentation der

unverdaulichen Kohlenhydrate entstehen (Cummings 1981). In Studien mit fettleibigen und dünnen Menschen wurde festgestellt, dass das Darmmikrobiom der dünnen Studienteilnehmergruppe einen signifikant höheren Anteil an *Bacteroidota* aufweist (Johnson et al. 2017). Die Lymphoma-Gruppe hat im Gegensatz zur gesunden Gruppe einen 8 % höheren Anteil an *Bacteroidota*, davon stammen nahezu alle aus der *Bacteroidaceae*-Familie. Es ist bekannt, dass die *Bacteroidaceae* durch die Aktivierung von CD4+T-Zellen, in die Immunantwort involviert ist. Individuen mit einem höheren Gehalt an *Bacteroidaceae* im Mikrobiom weisen eine höhere Konzentration an regulatorischen T-Zellen (Treg) und Myeloiden Suppressorzellen (MDSC) auf (Gopalakrishnan et al. 2018). Die Treg-Zellen unterdrücken abnormale oder übermäßige Immunantworten und hemmen somit auch die Antitumorimmunität. Eine hohe Treg-Zell Infiltration ist mit einer schlechten Prognose bei verschiedenen Krebsarten verbunden (Ohue und Nishikawa 2019). Im Gegensatz dazu sind die MDSC in der Lage sich bei Infektionen, Entzündungen oder Krebs auszudehnen und die T-Zellen vermittelte Immunantwort zu unterdrücken (Gabrilovich und Nagaraj 2009). Die bisherigen Ergebnisse stellen, durch eine aktive Beeinflussung der *Bacteroidaceae*-Konzentration im Mikrobiom, einen vielversprechenden Ansatz für die Immunotherapie bei Krebserkrankungen dar.

#### 4.5.3. *Fusobacteria*-assoziierte Bakterien

*Fusobacteria* sind ebenfalls gramnegative, obligat anaerobe Bakterien, welche bei Menschen und Tieren natürlicherweise im oberen Atmungstrakt und im Verdauungstrakt vorkommen (Allaker 2012). Die unterschiedlichen Spezies sind an einer Vielzahl von Infektionen beteiligt, in dem sie über geschädigte Schleimhautbarrieren in den Körper eindringen und sich dort vermehren (Booth 2014). In Humanstudien hat sich gezeigt, dass die Anzahl an *Fusobacteria*-assoziierten Spezies im Tumorgewebe von Darmkrebs, im Vergleich zu dem Gewebe der gesunden Darmschleimhaut deutlich erhöht ist (Kostic et al. 2012). In dieser Arbeit weisen *Fusobacteria* einen Anstieg von 12,34 % in der Lymphoma-Gruppe auf. Einige Spezies im Phylum der *Fusobacteria* können im Wirt proinflammatorische Reaktionen auslösen (Moore und Moore 1994). Sie besitzen spezifische Virulenzgene, die die Adhäsion an und das Eindringen in Epithelzellen ermöglichen (Han et al. 2000, Bachrach et al. 2005, Uitto et al. 2005). Ein vermehrtes Vorkommen dieser Bakterien in Tumorgewebe, kann ein Hinweis sein, dass sie zur Tumorentstehung beitragen, am wahrscheinlichsten durch einen Entzündungs-assoziierten Mechanismus (Kostic et al. 2012). Es ist aber auch möglich, dass sich diese Bakterien in der Tumor-Umgebung in späteren Tumor Stadien besser etablieren können als andere Bakterien, und damit keine Rolle in der Tumorentstehung haben. Es gibt noch keinen kausalen Zusammenhang zwischen *Fusobacteria* und Tumoren.

In den insgesamt 23 Lymphoma-Proben kann ein Anstieg von 12 % der Familie der *Fusobacteriaceae* beobachtet werden. Davon weisen 12 Proben einen Mittelwert von 25 % an *Fusobacteriaceae* und die anderen 11 Proben einen Mittelwert von 2 % aus. Die Familie der *Fusobacteriaceae* gelten als entzündungsfördernd und zeigen einen signifikanten Anstieg in Krebsgewebe (Sędzikowska und Szablewski 2021). Die großen Unterschiede innerhalb der Lymphoma-Gruppe können möglicherweise durch eine unterschiedliche Krankheitsdauer der Lymphoma betroffenen Katzen entstehen. Der tatsächliche Zeitpunkt, ab dem das Lymphom begonnen hat sich zu entwickeln, kann jedoch in der Praxis nicht festgestellt werden, da die Besitzer erst wesentlich später beim Auftreten gastrointestinaler Symptome wie Inappetenz, Erbrechen, Durchfall und Gewichtsabnahme vorstellig werden.

#### 4.5.4. *Proteobacteria*-assoziierte Bakterien

*Proteobacteria* sind gramnegative, fakultativ anaerobe Bakterien. Fakultativ anaerob bedeutet, dass die Bakterien in Abwesenheit und Anwesenheit von Sauerstoff vegetieren können. *Proteobacteria* weisen eine Vielzahl an Familien auf. Vermutlich können die Mitglieder der Familie *Enterobacteriaceae* besser unter entzündlichen Bedingungen im Darm überleben als die obligaten Anaerobier, welche für die Gesundheit des Darms verantwortlich sind. Dieses Phylum weist eine Vielzahl an pathogenen Arten gehörend z.B. zu *Campylobacter*, *Enterobacter* und *Escherichia* auf (Stecher et al. 2007). Die Daten dieser Arbeit ergeben insgesamt eine Vermehrung der *Proteobacteria* um 7,19 % in der Lymphoma-Gruppe. Auf Familienebene lässt sich ein Anstieg der *Enterobacteriaceae* von 2,89 % nachweisen, welche zu 95,5 % aus der *Echerichia-Shigella* Gruppe bestehen.

In Humanstudien konnte ein Zusammenhang von erhöhten *Burkholderiaceae*-Werten im neoplastischen Gewebe mit Darmkrebs hergestellt werden (Alomair et al. 2018). In der Lymphoma-Gruppe lässt sich ein Anstieg von 3 % der *Burkholderiaceae*-Familie erkennen. Diese Steigerung kann einen Auslöser für Neoplasien darstellen (Pederzoli et al. 2020).

#### 4.5.5. *Actinobacteria*-assoziierte Bakterien

*Actinobacteria* produzieren ebenso wie *Firmicutes* kurzkettige Fettsäuren und tragen damit zur Aufrechterhaltung der natürlichen Darmbarriere bei (Mondo et al. 2019). Das Phylum der *Actinobacteria*, speziell die Familie der *Bifidobacteriaceae*, weisen entzündungshemmende Eigenschaften über die Regulation von Immunzellen auf (O'Neill et al. 2017). Dabei können die Bakterien über bestimmte Oberflächenmoleküle die Immunzellen aktivieren. In vorangegangenen Studien zeigen Katzen mit Inflammatory Bowel Disease (IBD) und Small Cell Lymphoma (SCL) einen Rückgang der *Bifidobacteriaceae*. Diese sind direkt für die

Unterstützung der Darmbarriere verantwortlich (Marsilio et al. 2019). Die Lymphoma Katzen weisen eine Reduktion von 9,57 % des Phylum *Actinobacteria* auf. Innerhalb des Phylums *Actinobacteria* kommt es zu einer Abnahme der Familie der *Bifidobacteriaceae* um 1,22 %.

Die Familie der *Coriobacteriaceae* übernimmt wichtige Stoffwechselfunktionen, unter anderem die Umwandlung von Gallensalzen und Steroidhormonen (Clavel et al. 2014). Durch den Gallensäurestoffwechsel haben die Bakterien indirekt Auswirkung auf den Lipidstoffwechsel des Wirts. *Coriobacteriaceae* könnten ein Ansatz für das Erkennen von Stoffwechselstörungen darstellen, da sie mit einem gesunden Darmmikrobiom in Zusammenhang gebracht werden (Zhao et al. 2018). Die Lymphoma-Katzen weisen eine Reduktion von 7 % der *Coriobacteriaceae*-Familie auf. Davon liegen 7 Lymphoma-Katzen über dem Mittelwert von 2,5 % und die andern 13 Katzen darunter. Eine Reduktion dieser Bakterien in der Lymphoma-Gruppe könnte eine Veränderung des Stoffwechsels zur Folge haben oder durch die Stoffwechselveränderung die Lymphomerkrankung begünstigen.

#### 4.6. Bedeutung eines veränderten Mikrobioms der Lymphoma Katzen

„Eine schlechte Verdauung ist die Wurzel allen Übels“ (Hippokrates 460-370 v.Chr.). Die Darmschleimhaut ist die Schnittstelle unseres Körpers mit der Umwelt. Das Darmmikrobiom stellt eine wichtige Rolle in unserem Organismus dar. Es trägt zur Abwehr von pathogenen Bakterien bei und produziert Substanzen, die sowohl für die Bakterien als auch für den Wirt von Vorteil sind (Symbiose). Es fermentiert unverdauliche Kohlenhydrate und stellt SCFAs wie z.B. Propionat, Butyrat und Acetat her. Die Aufgabe der SCFAs ist die Senkung des intestinalen pH-Werts (Abwehrmechanismus) und die Unterstützung der epithelialen Darmzellen. Dysbiosen haben große Auswirkungen auf die Menge und die Zusammensetzung der Metaboliten, da diese zum Großteil von den Mikroben produziert werden. Zusätzlich hierzu spielen sie eine große Rolle bei Magen-Darm-, oder systemischen Erkrankungen (Walker und Lawley 2013).

Es gibt drei beschriebene Formen der Dysbiose. Erstens, eine Reduktion der Bakterienvielfalt; zweitens, ein Verlust der nützlichen Bakterien und drittens, die Zunahme von krankmachenden Bakterien (Mondo et al. 2019). Oft ist eine Dysbiose eine Mischform aus zwei bis drei dieser beschriebenen Zustände. Studien zum Vorhandensein von Dysbiosen bei Erkrankungen der Katze ergaben meist kein klar erkennbares Muster der Mikrobiomveränderungen (Mondo et al. 2019). Sehr oft kommt es unabhängig von der Art der Krankheit zu einer Verringerung der obligat anaeroben Bakterienvielfalt und zu einer Vermehrung der fakultativ anaeroben Bakterien wie den *Enterobacteriaceae* (Lupp et al. 2007, Stecher et al. 2007).

Untersuchungen ergaben, dass eine häufig beobachtete Veränderung des Mikrobioms bei Katzen mit IBD eine Abnahme von *Firmicutes* ist. Die Vermutung liegt nahe, weil *Firmicutes* eine wichtige Rolle in der Produktion von SCFAs spielt und damit die Darmgesundheit fördert (Suchodolski 2011b). Die Lymphoma Katzen weisen eine Reduktion von 17,87 % von *Firmicutes* auf. Dieses Phylum produziert vorwiegend Butyrat, was maßgeblich zur Aufrechterhaltung der Proliferation und Schleimproduktion der Epithelzellen im Darm beiträgt. Deren Ziel es ist die funktionelle Darmbarriere gegen mögliche pathogene Erreger aufrecht zu erhalten (Walker und Lawley 2013). Eine Abnahme der SCFAs produzierenden Bakterien in der Lymphoma-Gruppe verursacht möglicherweise eine Störung der Darmbarriere und erhöht damit die Anfälligkeit für potenzielle pathogene Erreger. Zusätzlich weisen die Lymphoma Katzen einen Anstieg von 8 % des Phylums *Bacteroidota* auf. Dieses Phänomen wurde bei einer Studie mit Mäusen, welche ein-drei Tage kein Futter bekamen, ebenfalls beobachtet. Die Kontrollgruppe an Mäusen, welche eine ad libitum Fütterung erhielt, wies einen signifikant geringeren Anteil an *Bacteroidota* auf (Kohl et al. 2014). Bei Katzen mit diagnostiziertem Lymphom konnte ein verminderter Appetit oder Anorexie bei 77,8% der Patienten festgestellt werden (Krick et al. 2008). Diese Ergebnisse lassen Grund zu der Annahme, dass Katzen der Lymphoma-Gruppe über längere Zeit der Erkrankung ein geringerer Anteil an kurzkettigen Kohlenhydraten zur Verfügung stehen.

Der Anstieg der *Fusobacteria* von 12,34 % in der Lymphoma-Gruppe lässt sich womöglich durch die bestehenden Lymphome erklären. Studien haben gezeigt, dass *Fusobacteria* gehäuft im Tumorgewebe auftreten (Kostic et al. 2012). Es sind weitere, standardisierte Versuchsmodelle nötig, um einen kausalen Zusammenhang zu prüfen. Die Lymphoma-Gruppe weist einen Anstieg der *Proteobacteria* und eine Zunahme der *Enterobacteriaceae* Familie, wovon der Großteil aus der Escherichia-Shigella Gruppe besteht, auf. Es wird angenommen, dass durch entzündliche Prozesse im Darm das Gleichgewicht der aeroben Darmschleimhaut und des anaeroben Darmlumens hin zu einem sauerstoffhaltigen Milieu gestört wird und somit die Vermehrung von *Enterobacteriaceae* fördert (Janeczko et al. 2008, Marsilio et al. 2019). Humanstudien haben gezeigt, dass diese Milieu das Tumorstadium beschleunigt (Zhou et al. 2021). Ebenso begünstigen die Veränderungen der Sauerstoffverhältnisse das Ansiedeln von fakultativ anaeroben pathogenen Arten. In der Lymphoma-Gruppe kommt es zu einer Abnahme des Phylums der *Actinobacteria*. In diesem Fall ist eine Störung der Darmbarriere einhergehend durch die Reduktion von SCFAs produzierenden und entzündungshemmenden Bakterien, vor allem aus der Familie der *Bifidobacteriaceae* denkbar. Der Zusammenhang eines gestörten Mikrobioms mit einer Krankheit und die genauen Wirkungsmechanismen sind weiterhin unklar (Lyu et al. 2020). Interessant wäre die zusätzliche Untersuchung der bakteriellen Stoffwechselprodukte (SCFs, Vitamine und Laktat (Macfarlane und Macfarlane 2003)) im Kot, um die Unterschiede des

veränderten Mikrobioms deutlicher darlegen zu können. Es ist bekannt, dass sich das Phylum *Bacteroidota* schnell an wechselnde Bedingungen anpassen kann (Johnson et al. 2017). Daher wären weitere Informationen über die Häufigkeit der Fütterung, das Futtermittel und die Menge, ein wichtiger Punkt in der Datenerfassung. Durch diese ergänzenden Informationen könnte eine Veränderung des Mikrobioms besser charakterisiert und die Aussagekraft erhöht werden.

#### 4.7. Zukunft der Mikrobiomanalyse

Die genaue taxonomische Zuordnung der repräsentativen Sequenzen ist von besonderer Bedeutung und ein sehr wichtiger Teil in der Mikrobiomanalyse (Allali et al. 2017).

In dieser Arbeit wurde mit ASVs gearbeitet. ASVs sind nicht geclusterte und bereits auf Fehler korrigierte Sequenzen. Innerhalb eines ASVs weisen die zugeordneten Sequenzen eine 100%ige Übereinstimmung auf, während bei der früher oft verwendeten OTU Klassifizierung nur eine 97%ige Ähnlichkeit gegeben ist. Der Vorteil von ASVs ist, dass diese eine höhere Spezifität und Sensitivität aufweisen. Somit können zum Beispiel ASVs, die von verschiedenen Studien oder Proben gewonnen wurden, miteinander verglichen werden oder ASV-Tabellen aus Projekten mit denselben Primerpaaren ergänzt werden. ASVs ermöglichen reproduzierbare, präzisere und wiederverwendbare Datensätze, weshalb die Verwendung von OTUs in der Mikrobiomforschung langfristig gesehen genauer hinterfragt werden sollte (Eren et al. 2015, Callahan et al. 2017). Sowohl in Mothur als auch in QIIME2 wird auf Grundlage der ASV-Tabelle die Datenanalyse (inklusive Beta- und Alpha Diversität) durchgeführt (Buza et al. 2019).

Die Zuverlässigkeit und die Integration von Metadaten wie z.B. der Gesundheitsstatus (Lymphoma/gesunde Gruppe) sind entscheidende Faktoren, um aufschlussreiche Erkenntnisse aus den Mikrobiomstudien zu erlangen. Diese wichtigen Informationen sind meist nicht öffentlich verfügbar, was zu Problemen bei der Wiederverarbeitung der Daten führt. Metadaten bilden den Schlüssel, um aus Daten biologische Erkenntnisse zu gewinnen (Moreno-Indias et al. 2021). Die Benutzer der Sequenzierungsplattformen sollten über die jeweiligen Schwachpunkte der Programme Bescheid wissen, um dies bei der Auswahl berücksichtigen zu können. QIIME2 bietet im Moment als einzige Analyseplattform die Möglichkeit, Datensätze retrospektiv nachzuverfolgen und damit die Reproduzierbarkeit sicherzustellen. In Zukunft sollten im Studiendesign metagenomische Ansätze, die die Gesamtheit aller mikrobiellen Gene berücksichtigen, sowie Metatranskriptom Analysen, die die Genexpression von Mikroben in natürlichen Umgebungen untersuchen, forciert werden, um funktionelle Mikrogen-Wirt Interaktionen untersuchen zu können.

## 4.8. Ausblick

Die Mikrobiomauswertung ist sowohl in der Datenanalyse als auch in der anschließenden Auswertung und Interpretation der Ergebnisse herausfordernd. Studien konzentrieren sich häufig nur auf die Veränderungen des Mikrobioms bei Pathologien des Darmtrakts und nicht darauf, wie diese Dysbiose wieder in Richtung des gesunden Gleichgewichts (Eubiose) verschoben werden kann und was die Auswirkungen auf die Grunderkrankung wären. Das Mikrobiom kann bei Menschen durch Ernährung, antimikrobielle Behandlung, Probiotika und Präbiotika reguliert werden und dadurch die Krebsbehandlung positiv beeinflussen (Zhou et al. 2021). Man hat herausgefunden, dass der Nahrung zugesetzte lösliche Ballaststoffquellen, sogenannte Präbiotika, und der Gehalt an Mikronährstoffe (z.B. Proteingehalt) im Futter mit einer Veränderung des fäkalen Mikrobioms bei gesunden Katzen in Zusammenhang gebracht werden kann. Des Weiteren könnte dies eine positive Auswirkung auf Tiere mit chronischen Darmentzündungen haben (Vester et al. 2009, Barry et al. 2010). Die Verabreichung von bestimmten Bakterienstämmen hätte somit Potential, die Darmmikrobiota von chronischen Magen-Darm-Erkrankungen positiv zu verändern oder sogar Dysbiosen auszugleichen (Hart et al. 2012). Hier wäre es von Vorteil, weiter an der genauen Erklärung der Wirkungsmechanismen zu arbeiten.

In dieser Arbeit zeigen nicht allen Proben der Lymphoma Katzen einen Unterschied zu der Kontrollgruppe. Manche Proben clustern mit den der gesunden Gruppe, andere Proben sind diffus verteilt. Daher kann hier kein klares „Lymphoma-Muster“ erkannt werden, welches repräsentativ für die Lymphomerkrankung wäre. Aufgrund dessen ist es aktuell noch nicht möglich, einen oder mehrere Biomarker zu definieren. Wäre dies in Zukunft der Fall, hätte das in der Tiermedizin einen enormen Vorteil, sowohl für die Tiere als auch für die Besitzer und die Tierärzte. Die Chance, Krankheiten noch vor Symptombeginn zu erkennen und dadurch die Wahrscheinlichkeit, eine bessere Prognose zu erlangen, ist vielversprechend. Beim Menschen wurden bereits signifikante Veränderungen in dem Darmmikrobiom von Patienten mit verschiedenen Krebsarten gefunden, dabei dienen spezifische Bakterien oder spezielle Mikrobiom-Eigenschaften als Marker für Screenings und Prognosevorhersagen (Zhou et al. 2021). Katzen sind nicht nur im Bereich der Mikrobiomforschung ein interessantes Forschungsmodell, sondern lassen auch Rückschlüsse auf menschliche Pathologien wie Krebs, Typ-2-Diabetes mellitus oder HIV (Humanes Immundefizienz-Virus) zu (McNiel 2001, Henson und O'Brien 2006, Bienzle 2014). Dennoch ist das Wissen über das Mikrobiom der Katzen so gering, dass es noch keinen Nutzen für die Praxis aufweist (Lyu et al. 2020).

Der enorme Arbeits- und Kostenaufwand dieser Studien könnte durch die Weiterentwicklung von NGS und Automatisierung von bioinformatischer Analysetools wie z.B. durch QIIME2 minimiert werden (Moreno-Indias et al. 2021). Von daher ist folgender Grundsatz für die Entwicklung von bioinformatischen Analysetools in der Mikrobiomforschung von großer

Bedeutung: FAIR (Findable, Accessible, Interoperable and Reusable) (Shetty und Lahti 2019), was zu Deutsch „Auffindbarkeit, Zugänglichkeit, vollständige Kompatibilität und Wiederverwendbarkeit“ bedeutet. Derzeit gib es auf dem Markt kein frei zugängliches bioinformatisches Analysetool, das den Anforderungen der Mikrobiomanalyse so sehr entspricht wie QIIME2.

## 5. Zusammenfassung

Das Mikrobiom der Katzen beinhaltet eine Vielzahl an Mikroorganismen, die sich je nach Fütterung, Haltungsbedingungen, Alter und Gesundheitsstatus anpasst. Ziel dieser Arbeit war, die bioinformatische Analyse eines Probensatzes bestehend aus Kot von an Lymphoma erkrankten Katzen und von einer gesunden Kontrollgruppe. Die bioinformatische Auswertung wurde mit dem Tool QIIME2 gemacht. Im Mittelpunkt der Analyse stand die Anwendung eines Workflows und Interpretation des mikrobiellen Datensatzes. Dazu wurde aus 33 felinen Proben (23 von Lymphoma Katzen, 10 von gesunden Katzen), und einer Leerkontrolle DNA extrahiert. Anschließend wurde das 16S rRNA Gen amplifiziert und mittels Illumina MiSeq sequenziert. Die Datenanalyse in QIIME2 ergab, dass 11.006.277 Sequenzen generiert wurden und nach der Qualitätskontrolle 65,31 % der Sequenzen verblieben. Es wurden 3005 ASVs (Amplicon Sequence Variants) generiert, welche insgesamt 16 Phyla zugeordnet werden konnten. Die am häufigsten vorkommenden Phyla waren *Firmicutes* (47,99 %), *Bacteroidota* (28,01 %), *Fusobacteria* (13,64 %), *Proteobacteria* (6,90 %) und *Actinobacteria* (2,51 %). Die Mikrobiomstruktur der beiden Gruppen (Lymphoma/Gesund) zeigte einen signifikanten Unterschied ( $p=0,003$ ). Die Tiere der gesunden Gruppe wiesen einen höheren Gehalt an *Firmicutes* (1,42 fold change), *Actinobacteria* (4,43 fold change) und einen niedrigeren Anteil an *Bacteroidota* (0,74 fold change), *Proteobacteria* (0,21 fold change) und *Fusobacteria* (0,13 fold change) im Gegensatz zu der Lymphoma-Gruppe auf. Die Veränderungen des Mikrobioms in der Lymphoma-Gruppe haben negative Auswirkungen auf den Schleimhautschutz und die Aufrechterhaltung der natürlichen Darmbarriere vor Pathogenen. Die Anfälligkeit für Entzündungen steigt einerseits, durch die Reduktion von *Firmicutes*-assoziierten Bakterien, welche zur Aufrechterhaltung der Darmhomöostase und wichtigen Stoffwechselfunktionen des Wirts beitragen. Andererseits durch eine Reduktion der *Fusobacteria*-assoziierten Spezies, die eine proinflammatorische Wirkung auf das Tumorgewebe von Darmkrebs aufweisen. Zusätzlich haben *Bacteroidota*-assoziierten Bakterien Auswirkungen auf das Immunsystem, reduzieren die Antitumorantwort und kommen vermehrt in der Lymphoma-Gruppe vor. Zusammenfassend werden in dieser Arbeit erstmals mikrobielle Shifts im Kot von Lymphomakatten beschrieben.

## 6. Summary

The microbiome of cats contains a variety of microorganisms that adapt according to feeding, housing conditions, age and health status. The aim of this thesis was the bioinformatic analysis of a sample set of feces from cats suffering from lymphoma and from a healthy control group. The bioinformatic analysis was done with the tool QIIME2. The focus of the analysis was the elaboration of a workflow and the identification of strengths and weaknesses of the bioinformatic tool (QIIME2). For this purpose, DNA was extracted from 33 feline samples from lymphoma cats, 10 from healthy cats) and a blank control. Subsequently, the 16S rRNA gene was amplified and sequenced using Illumina MiSeq. Data analysis in QIIME2 showed that 11,006,277 sequences were generated and 65.31% of the sequences remained after quality control. There were 3005 ASVs (Amplicon Sequence Variants) generated, which could be assigned to a total of 16 phyla. The most abundant phyla were *Firmicutes* (47,99 %), *Bacteroidota* (28,01 %), *Fusobacteria* (13,64 %), *Proteobacteria* (6,90 %), and *Actinobacteria* (2,51 %). The microbiota structure of the two groups (lymphoma/healthy) showed a significant difference ( $p=0.003$ ). The cats of the healthy group had a higher content of *Firmicutes* (1,42-fold change), *Actinobacteria* (4,43-fold change) and a lower percentage of *Bacteroidota* (0,74-fold change), *Proteobacteria* (0,21-fold change) and *Fusobacteria* (0,13-fold change) in contrast to the lymphoma group. The alterations of the microbiome in the lymphoma group has negative effects on mucosal protection and maintenance of the intestinal barrier which is important to fend off pathogens. The susceptibility to inflammation increases 1. due to the reduction of *Firmicutes*-associated bacteria, which contribute to the maintenance of intestinal homeostasis and important metabolic functions of the host and 2. Due to the increase of *Fusobacteria*-associated species, which have a proinflammatory effect on colon cancer tumor tissue. In addition, *Bacteroidota*-associated bacteria have effects on the immune system, reduce the antitumor response and occur increasing in the lymphoma group. In summary, this work describes microbial shifts in faeces of lymphoma cats for the first time.

## 7. Abkürzungsverzeichnis

dNTP	Desoxyribonukleosid-Triphosphate
ddNTP	Didesoxyribonucleosid-Triphosphate
FSG	First-Generation-Sequencing
SGS	Second-Generation-Sequencing
TGS	Third-Generation-Sequencing
SBS	Sequencing by Synthesis
SBL	Sequencing by Hybridization and Ligation
SOLiD	Support Oligonucleotide Ligation Detection
SMRT	Single Molecule Real Time
NCBI	National Center for Biotechnology Information
PEAR	Paired-End reAd mergeR
QIIME	Quantitative Insights Into Microbial Ecology
DADA	Divisive Amplicon Denoising Algorithm
MICCA	MICRobial Community Analyse
PICRUSt2	Phylogenetic Investigation of Communities by Reconstruction of Unobserved States
GIT	Gastrointestinaltrakt
16S rRNA	16S-Untereinheit der rRNA von Prokaryoten
Abb.	Abbildung
Tab.	Tabelle
bp	Basenpaare
RNA	Ribonukleinacid
DNA	Desoxyribonukleinacid
ASV	Amplicon Sequence Variant
OUT	Operative taxonomische Einheit (Operational Taxonomic Unit)
PCR	Polymerase-Kettenreaktion (Polymerase Chain Reaction)
qPCR	real time Polymerase-Kettenreaktion
dPCR	droplet digital PCR
NGS	Next-Generation-Sequencing
SCFA	short-chain-fatty-acids
IBD	Inflammatory-bowel-disease
MDSC	Myeloiden Suppressorzellen (myeloide suppressorcells)

## 8. Literaturverzeichnis

Albanese D, Fontana P, De Filippo C, Cavalieri D, Donati C. 2015. MICCA: a complete and accurate software for taxonomic profiling of metagenomic data. *Scientific Reports*, 5:9743.

Allaker RP. 2012. 36 - Non-sporing anaerobes: Wound infection; periodontal disease; abscess; normal flora. In: Greenwood D, Barer M, Slack R, Irving W, Hrsg. *Medical Microbiology* (Eighteenth Edition). 359–364, Edinburgh: Churchill Livingstone;

Allali I, Arnold JW, Roach J, Cadenas MB, Butz N, Hassan HM, Koci M, Ballou A, Mendoza M, Ali R, et al. 2017. A comparison of sequencing platforms and bioinformatics pipelines for compositional analysis of the gut microbiome. *BMC Microbiology*, 17 DOI:10.1186/s12866-017-1101-8

Alomair AO, Masoodi I, Alyamani EJ, Allehibi AA, Qutub AN, Alsayari KN, Altammami MA, Alsharqeti AS. 2018. Colonic Mucosal Microbiota in Colorectal Cancer: A Single-Center Metagenomic Study in Saudi Arabia. *Gastroenterology Research and Practice*, 2018:5284754.

Ambardar S, Gupta R, Trakroo D, Lal R, Vakhlu J. 2016. High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian Journal of Microbiology*, 56(4):394–404.

Austin B. 2017. The value of cultures to modern microbiology. *Antonie van Leeuwenhoek*, 110(10):1247–1256.

Ayrapetyan M, Oliver JD. 2016. The viable but non-culturable state and its relevance in food safety. *Current Opinion in Food Science*, 8:127–133.

Bachrach G, Ianculovici C, Naor R, Weiss EI. 2005. Fluorescence based measurements of *Fusobacterium nucleatum coaggregation* and of fusobacterial attachment to mammalian cells. *FEMS microbiology letters*, 248(2):235–240.

Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. 2005. Host-bacterial mutualism in the human intestine. *Science (New York, N.Y.)*, 307(5717):1915–1920.

Barry KA, Wojcicki BJ, Middelbos IS, Vester BM, Swanson KS, Fahey GC. 2010. Dietary cellulose, fructooligosaccharides, and pectin modify fecal protein catabolites and microbial populations in adult cats. *Journal of Animal Science*, 88(9):2978–2987.

Barry KA, Middelbos IS, Vester Boler BM, Dowd SE, Suchodolski JS, Henrissat B, Coutinho PM, White BA, Fahey GC, Swanson KS. 2012. Effects of dietary fiber on the feline gastrointestinal metagenome. *Journal of Proteome Research*, 11(12):5924–5933.

Bienzle D. 2014. FIV in cats--a useful model of HIV in people? *Veterinary Immunology and Immunopathology*, 159(3–4):171–179.

Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Gregory Caporaso J. 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*, 6(1):90.

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, et al. 2019a. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature biotechnology*, 37(8):852–857.

Book M, Lehmann LE, Zhang XH, Stueber F. 2009. *Real-Time PCR in Microbiology: From Diagnosis to*

Characterization. In: Rello J, Díaz E, Rodríguez A, Hrsg. Management of Sepsis: The PIRO Approach. 65–85, Berlin, Heidelberg: Springer Berlin Heidelberg;

Buza TM, Tonui T, Stomeo F, Tiambo C, Katani R, Schilling M, Lyimo B, Gwakisa P, Cattadori IM, Buza J, et al. 2019. iMAP: an integrated bioinformatics and visualization pipeline for microbiome data analysis. BMC Bioinformatics, 20 DOI:10.1186/s12859-019-2965-4

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods, 13(7):581–583.

Callahan BJ, McMurdie PJ, Holmes SP. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME Journal, 11(12):2639–2643.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, et al. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods, 7(5):335–336.

Chang JY, Antonopoulos DA, Kalra A, Tonelli A, Khalife WT, Schmidt TM, Young VB. 2008. Decreased Diversity of the Fecal Microbiome in Recurrent *Clostridium difficile* –Associated Diarrhea. The Journal of Infectious Diseases, 197(3):435–438.

Chao A. 1984. Nonparametric Estimation of the Number of Classes in a Population. Scandinavian Journal of Statistics, 11(4):265–270.

Chen L, Liu P, Evans TC, Ettwiller LM. 2017. DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. Science (New York, N.Y.), 355(6326):752–756.

Chen W, Liu F, Ling Z, Tong X, Xiang C. 2012. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. PloS One, 7(6):e39743.

Clavel T, Lepage P, Charrier C. 2014. The Family *Coriobacteriaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, Hrsg. The Prokaryotes: *Actinobacteria*. 201–238, Berlin, Heidelberg: Springer.

Comalada M, Bailón E, de Haro O, Lara-Villoslada F, Xaus J, Zarzuelo A, Gálvez J. 2006. The effects of short-chain fatty acids on colon epithelial proliferation and survival depend on the cellular phenotype. Journal of Cancer Research and Clinical Oncology, 132(8):487–497.

Cummings JH. 1981. Short chain fatty acids in the human colon. Gut, 22(9):763–779.

Deng P, Swanson KS. 2015. Gut microbiota of humans, dogs and cats: current knowledge and future opportunities and challenges. British Journal of Nutrition, 113(S1):S6–S17.

Desai AR, Musil KM, Carr AP, Hill JE. 2009. Characterization and quantification of feline fecal microbiota using cpn60 sequence-based methods and investigation of animal-to-animal variation in microbial population structure. Veterinary Microbiology, 137(1–2):120–128.

Deusch O, O'Flynn C, Colyer A, Swanson KS, Allaway D, Morris P. 2015. A Longitudinal Study of the Feline Faecal Microbiome Identifies Changes into Early Adulthood Irrespective of Sexual Development. PloS One, 10(12):e0144881.

van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. 2014. Ten years of next-generation sequencing technology. Trends in genetics: TIG, 30(9):418–426.

Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, Huttenhower C, Langille MGI. 2020. PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology*, 38(6):685–688.

Dunlap PV. 2001. Microbial Diversity. In: Levin SA, Hrsg. *Encyclopedia of Biodiversity* (Second Edition). 280–291, Waltham: Academic Press.

Eisenstein M. 2020. The skin microbiome. *Nature*, 588(7838):S209–S209.

Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML. 2015. Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *The ISME journal*, 9(4):968–979.

Escobar-Zepeda A, Vera-Ponce de León A, Sanchez-Flores A. 2015. The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics. *Frontiers in Genetics*, 6 DOI:10.3389/fgene.2015.00348.

Estaki M, Jiang L, Bokulich NA, McDonald D, González A, Kosciolk T, Martino C, Zhu Q, Birmingham A, Vázquez-Baeza Y, et al. 2020. QIIME 2 Enables Comprehensive End-to-End Analysis of Diverse Microbiome Data and Comparative Studies with Publicly Available Data. *Current Protocols in Bioinformatics*, 70(1):e100.

Fei N, Zhao L. 2013. An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice. *The ISME journal*, 7(4):880–884.

Fournier P-E, Drancourt M, Raoult D. 2007. Bacterial genome sequencing and its use in infectious diseases. *The Lancet Infectious Diseases*, 7(11):711–723.

Frank DN, Amand ALS, Feldman RA, Boedeker EC, Harpaz N, Pace NR. 2007. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences*, 104(34):13780–13785.

Funari V, Canosa SJ. 2014. The Importance of Bioinformatics in NGS: Breaking the Bottleneck in Data Interpretation. *Science*, 344(6184):653–653.

Gabrilovich DI, Nagaraj S. 2009. Myeloid-derived suppressor cells as regulators of the immune system. *Nature Reviews Immunology*, 9(3):162–174.

Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpinets TV, Prieto PA, Vicente D, Hoffman K, Wei SC, et al. 2018. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science (New York, N.Y.)*, 359(6371):97–103.

Haines JL, Pericak-Vance MA. 2007. *Genetic Analysis of Complex Disease*. John Wiley & Sons.  
Hall M, Beiko RG. 2018. 16S rRNA Gene Analysis with QIIME2. In: Beiko RG, Hsiao W, Parkinson J, Hrsg. *Microbiome Analysis: Methods and Protocols*. 113–129, New York, NY: Springer.

Han YW, Shi W, Huang GT, Kinder Haake S, Park NH, Kuramitsu H, Genco RJ. 2000. Interactions between periodontal bacteria and human oral epithelial cells: *Fusobacterium nucleatum* adheres to and invades epithelial cells. *Infection and Immunity*, 68(6):3140–3146.

Handl S, Dowd SE, Garcia-Mazcorro JF, Steiner JM, Suchodolski JS. 2011. Massive parallel 16S rRNA gene pyrosequencing reveals highly diverse fecal bacterial and fungal communities in healthy dogs and cats: Fecal microbiota in dogs and cats using pyrosequencing. *FEMS Microbiology Ecology*, 76(2):301–310.

- Hart ML, Suchodolski JS, Steiner JM, Webb CB. 2012. Open-label trial of a multi-strain synbiotic in cats with chronic diarrhea. *Journal of Feline Medicine and Surgery*, 14(4):240–245.
- Heather JM, Chain B. 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107(1):1–8.
- Henson MS, O'Brien TD. 2006. Feline models of type 2 diabetes mellitus. *ILAR journal*, 47(3):234–242. Human Genome Project FAQ. [Genome.gov](http://Genome.gov).
- James Booth S. 2014. *Fusobacterium Infections*. Reference Module in Biomedical Sciences. Elsevier; DOI:10.1016/B978-0-12-801238-3.05030-3
- Janeczko S, Atwater D, Bogel E, Greiter-Wilke A, Gerold A, Baumgart M, Bender H, McDonough PL, McDonough SP, Goldstein RE, et al. 2008. The relationship of mucosal bacteria to duodenal histopathology, cytokine mRNA, and clinical disease activity in cats with inflammatory bowel disease. *Veterinary Microbiology*, 128(1–2):178–193.
- Johnson EL, Heaver SL, Walters WA, Ley RE. 2017. Microbiome and metabolic disease: revisiting the bacterial phylum *Bacteroidetes*. *Journal of Molecular Medicine (Berlin, Germany)*, 95(1):1–8.
- Johnston K, Lamport A, Batt RM. 1993. An unexpected bacterial flora in the proximal small intestine of normal cats. *The Veterinary Record*, 132(14):362–363.
- Kaakoush NO. 2015. Insights into the Role of Erysipelotrichaceae in the Human Host. *Frontiers in Cellular and Infection Microbiology*, 5:84.
- Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, 41(1):e1.
- Kohl KD, Amaya J, Passemont CA, Dearing MD, McCue MD. 2014. Unique and shared responses of the gut microbiota to prolonged fasting: a comparative study across five classes of vertebrate hosts. *FEMS microbiology ecology*, 90(3):883–894.
- Kolmeder CA, Ritari J, Verdam FJ, Muth T, Keskitalo S, Varjosalo M, Fuentes S, Greve JW, Buurman WA, Reichl U, et al. 2015. Colonic metaproteomic signatures of active bacteria and the host in obesity. *Proteomics*, 15(20):3544–3552.
- Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, Ojesina AI, Jung J, Bass AJ, Tabernero J, et al. 2012. Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome Research*, 22(2):292–298.
- Krick EL, Little L, Patel R, Shofer FS, Sorenmo K, Clifford CA, Baez JL. 2008. Description of clinical and pathological findings, treatment and outcome of feline large granular lymphocyte lymphoma (1996–2004). *Veterinary and Comparative Oncology*, 6(2):102–110.
- Kuczynski J, Stombaugh J, Walters WA, González A, Caporaso JG, Knight R. 2011. Using QIIME to analyze 16S rRNA gene sequences from Microbial Communities. *Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.]*, Chapter:Unit10.7 DOI: 10.1002/0471250953.bi1007s36.
- Lagier J-C, Edouard S, Pagnier I, Mediannikov O, Drancourt M, Raoult D. 2015. Current and Past Strategies for Bacterial Culture in Clinical Microbiology. *Clinical Microbiology Reviews*, 28(1):208–236.

- Lin R, Liu W, Piao M, Zhu H. 2017. A review of the relationship between the gut microbiota and amino acid metabolism. *Amino Acids*, 49(12):2083–2090.
- Liu C, Cui Y, Li X, Yao M. 2021. *microeco*: an R package for data mining in microbial community ecology. *FEMS Microbiology Ecology*, 97(2):255.
- Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L. 2018. Phylogenetically Novel Uncultured Microbial Cells Dominate Earth Microbiomes. *mSystems*, 3(5).
- Locey KJ, Lennon JT. 2016. Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences*, 113(21):5970–5975.
- Logares R, Haverkamp THA, Kumar S, Lanzén A, Nederbragt AJ, Quince C, Kauserud H. 2012. Environmental microbiology through the lens of high-throughput DNA sequencing: synopsis of current platforms and bioinformatics approaches. *Journal of Microbiological Methods*, 91(1):106–113.
- Lombardo M. 2008. Access to mutualistic endosymbiotic microbes: An underappreciated benefit of group living. *Behavioral Ecology and Sociobiology*, 62(4), 479–497
- López-García A, Pineda-Quiroga C, Atxaerandio R, Pérez A, Hernández I, García-Rodríguez A, González-Recio O. 2018. Comparison of Mothur and QIIME for the Analysis of Rumen Microbiota Composition Based on 16S rRNA Amplicon Sequences. *Frontiers in Microbiology*, 9:3010.
- Lozupone CA, Hamady M, Kelley ST, Knight R. 2007. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Applied and Environmental Microbiology*, 73(5):1576–1585.
- Lupp C, Robertson ML, Wickham ME, Sekirov I, Champion OL, Gaynor EC, Finlay BB. 2007. Host-Mediated Inflammation Disrupts the Intestinal Microbiota and Promotes the Overgrowth of Enterobacteriaceae. *Cell Host & Microbe*, 2(2):119–129.
- Lyu Y, Su C, Verbrugghe A, Van de Wiele T, Martos Martinez-Caja A, Hesta M. 2020. Past, Present, and Future of Gastrointestinal Microbiota Research in Cats. *Frontiers in Microbiology*, 11:1661.
- Macfarlane S, Macfarlane GT. 2003. Regulation of short-chain fatty acid production. *The Proceedings of the Nutrition Society*, 62(1):67–72.
- Madigan MT. 2015. *Brock biology of microorganisms*. Harlow: Pearson Education ISBN: 978-0-321-64963-8.
- Mai V, Colbert LH, Perkins SN, Schatzkin A, Hursting SD. 2007. Intestinal microbiota: a potential diet-responsive prevention target in *ApcMin* mice. *Molecular Carcinogenesis*, 46(1):42–48.
- Major G, Spiller R. 2014. Irritable bowel syndrome, inflammatory bowel disease and the microbiome. *Current Opinion in Endocrinology, Diabetes, and Obesity*, 21(1):15–21.
- Marsilio S, Pilla R, Sarawichitr B, Chow B, Hill SL, Ackermann MR, Estep JS, Lidbury JA, Steiner JM, Suchodolski JS. 2019. Characterization of the fecal microbiome in cats with inflammatory bowel disease or alimentary small cell lymphoma. *Scientific Reports*, 9 DOI: 10.1038/s41598-019-55691-w.
- McNiel EA. 2001. Vaccine-associated sarcomas in cats: a unique cancer model. *Clinical Orthopaedics and Related Research*, (382):21–27.
- Meazzi S, Stranieri A, Lauzi S, Bonsembiante F, Ferro S, Paltrinieri S, Giordano A. 2019. Feline gut

- microbiota composition in association with feline coronavirus infection: A pilot study. *Research in Veterinary Science*, 125:272–278.
- Metzker ML. 2005. Emerging technologies in DNA sequencing. *Genome Research*, 15(12):1767–1776.
- Metzker ML. 2010. Sequencing technologies — the next generation. *Nature Reviews Genetics*, 11(1):31–46.
- Minamoto Y, Hooda S, Swanson KS, Suchodolski JS. 2012a. Feline gastrointestinal microbiota. *Animal Health Research Reviews*, 13(1):64–77.
- Mondo E, Marliani G, Accorsi PA, Cocchi M, Di Leone A. 2019. Role of gut microbiota in dog and cat's health and diseases. *Open Veterinary Journal*, 9(3):253–258.
- Moon CD, Young W, Maclean PH, Cookson AL, Bermingham EN. 2018. Metagenomic insights into the roles of Proteobacteria in the gastrointestinal microbiomes of healthy dogs and cats. *MicrobiologyOpen*, 7(5).
- Moore JH. 2007. Bioinformatics. *Journal of Cellular Physiology*, 213(2):365–369.
- Moore WE, Moore LV. 1994. The bacteria of periodontal diseases. *Periodontology 2000*, 5:66–77.
- Moreno-Indias I, Lahti L, Nedyalkova M, Elbere I, Roshchupkin G, Adilovic M, Aydemir O, Bakir-Gungor B, Santa Pau EC, D'Elia D, et al. 2021. Statistical and Machine Learning Techniques in Human Microbiome Studies: Contemporary Challenges and Solutions. *Frontiers in Microbiology*, 12 DOI: 10.3389/fmicb.2021.635781.
- Mortarino M, Franceschi A, Mancianti F, Bazzocchi C, Genchi C, Bandi C. 2004. Quantitative PCR in the diagnosis of *Leishmania*. *Parassitologia*, 46(1–2):163–167.
- Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, 51 Pt 1:263–273.
- Nagendra H. 2002. Opposite trends in response for the Shannon and Simpson indices of landscape diversity. *Applied Geography*, 22(2):175–186.
- Neish AS. 2009. Microbes in Gastrointestinal Health and Disease. *Gastroenterology*, 136(1):65–80.
- Ogilvie LA, Jones BV. 2012. Dysbiosis modulates capacity for bile acid modification in the gut microbiomes of patients with inflammatory bowel disease: a mechanism and marker of disease? *Gut*, 61(11):1642–1643.
- Ohue Y, Nishikawa H. 2019. Regulatory T (Treg) cells in cancer: Can Treg cells be a new therapeutic target? *Cancer Science*, 110(7):2080–2089.
- O'Neill I, Schofield Z, Hall LJ. 2017. Exploring the role of the microbiota member *Bifidobacterium* in modulating immune-linked diseases. *Emerging Topics in Life Sciences*, 1(4):333–349.
- Orchel A, Dzierzewicz Z, Parfiniewicz B, Weglarz L, Wilczok T. 2005. Butyrate-induced differentiation of colon cancer cells is PKC and JNK dependent. *Digestive Diseases and Sciences*, 50(3):490–498.
- Palm NW, de Zoete MR, Cullen TW, Barry NA, Stefanowski J, Hao L, Degnan PH, Hu J, Peter I, Zhang W, et al. 2014. Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease.

Cell, 158(5):1000–1010.

Papasouliotis K, Sparkes AH, Werrett G, Egan K, Gruffydd-Jones EA, Gruffydd-Jones TJ. 1998. Assessment of the bacterial flora of the proximal part of the small intestine in healthy cats, and the effect of sample collection method. *American Journal of Veterinary Research*, 59(1):48–51.

Pareek CS, Smoczynski R, Tretyn A. 2011. Sequencing technologies and genome sequencing. *Journal of Applied Genetics*, 52(4):413–435.

Pederzoli F, Ferrarese R, Amato V, Locatelli I, Alchera E, Lucianò R, Nebuloni M, Briganti A, Gallina A, Colombo R, et al. 2020. Sex-specific Alterations in the Urinary and Tissue Microbiome in Therapy-naïve Urothelial Bladder Cancer Patients. *European Urology Oncology*, 3(6):784–788.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. 2011. Scikit-learn: Machine Learning in Python. *Machine learning in Python*, 6.

Pichler M, Coskun ÖK, Ortega-Arbulú A, Conci N, Wörheide G, Vargas S, Orsi WD. 2018. A 16S rRNA gene sequencing and analysis protocol for the Illumina MiniSeq platform. *MicrobiologyOpen*, 7(6) DOI:10.1002/mbo3.611.

Pielou EC. 1966. The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology*, 13:131–144.

Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, 13(1):1–13.

Ramotar K, Conly JM, Chubb H, Louie TJ. 1984. Production of menaquinones by intestinal anaerobes. *The Journal of Infectious Diseases*, 150(2):213–218.

Rhoads A, Au KF. 2015. PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics*, 13(5):278–289.

Ritchie LE, Steiner JM, Suchodolski JS. 2008. Assessment of microbial diversity along the feline intestinal tract using 16S rRNA gene analysis. *FEMS Microbiology Ecology*, 66(3):590–598.

Sanford JA, Gallo RL. 2013. Functions of the skin microbiota in health and disease. *Seminars in Immunology*, 25(5):370–377.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12):5463–5467.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, et al. 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*, 75(23):7537–7541.

Sędzikowska A, Szablewski L. 2021. Human Gut Microbiota in Health and Selected Cancers. *International Journal of Molecular Sciences*, 22(24):13440.

Shetty SA, Lahti L. 2019. Microbiome data science. *Journal of Biosciences*, 44(5).

Simpson EH. 1949. Measurement of Diversity. *Nature*, 163(4148):688–688.

- Singh S, Eldin C, Kowalczywska M, Raoult D. 2013. Axenic culture of fastidious and intracellular bacteria. *Trends in Microbiology*, 21(2):92–99.
- Sinha SR, Haileselassie Y, Nguyen LP, Tropini C, Wang M, Becker LS, Sim D, Jarr K, Spear ET, Singh G, et al. 2020. Dysbiosis-induced Secondary Bile Acid Deficiency Promotes Intestinal Inflammation. *Cell host & microbe*, 27(4):659-670.e5.
- Slatko BE, Gardner AF, Ausubel FM. 2018. Overview of Next Generation Sequencing Technologies. *Current protocols in molecular biology*, 122(1):e59.
- Stecher B, Robbiani R, Walker AW, Westendorf AM, Barthel M, Kremer M, Chaffron S, Macpherson AJ, Buer J, Parkhill J, et al. 2007. Salmonella enterica Serovar Typhimurium Exploits Inflammation to Compete with the Intestinal Microbiota. *PLOS Biology*, 5(10):e244.
- Suchodolski JS. 2011a. Intestinal Microbiota of Dogs and Cats: a Bigger World than We Thought. *The Veterinary Clinics of North America. Small Animal Practice*, 41(2):261–272.
- Suchodolski JS. 2011b. Companion Animal Symposium: Microbes and gastrointestinal health of dogs and cats. *Journal of Animal Science*, 89(5):1520–1530.
- Suchodolski JS, Dowd SE, Westermarck E, Steiner JM, Wolcott RD, Spillmann T, Harmoinen JA. 2009. The effect of the macrolide antibiotic tylosin on microbial diversity in the canine small intestine as demonstrated by massive parallel 16S rRNA gene sequencing. *BMC microbiology*, 9:210.
- Suchodolski JS, Foster ML, Sohail MU, Leutenegger C, Queen EV, Steiner JM, Marks SL. 2015. The Fecal Microbiome in Cats with Diarrhea. *PLOS ONE*, 10(5):e0127378.
- Tan HT, Tan S, Lin Q, Lim TK, Hew CL, Chung MCM. 2008. Quantitative and temporal proteome analysis of butyrate-treated colorectal cancer cells. *Molecular & cellular proteomics: MCP*, 7(6):1174–1185.
- Theil S, Rifa E. 2021. rANOMALY: AmplicoN wOrkflow for Microbial community AnaLYsis. *F1000Research*, 10:7.
- Thürmer A. 2014. Next Generation Sequencing in der mikrobiellen (Meta)Genomforschung: Genomsequenzierung. *BIOspektrum*, 20(2):168–171.
- Tremaroli V, Bäckhed F. 2012. Functional interactions between the gut microbiota and host metabolism. *Nature*, 489(7415):242–249.
- Tun HM, Brar MS, Khin N, Jun L, Hui RK-H, Dowd SE, Leung FC-C. 2012. Gene-centric metagenomics analysis of feline intestinal microbiome using 454 junior pyrosequencing. *Journal of Microbiological Methods*, 88(3):369–376.
- Uitto V-J, Baillie D, Wu Q, Gendron R, Grenier D, Putnins EE, Kanervo A, Firth JD. 2005. *Fusobacterium nucleatum* increases collagenase 3 production and migration of epithelial cells. *Infection and Immunity*, 73(2):1171–1179.
- Vasar M, Davison J, Neuenkamp L, Sepp S-K, Young JPW, Moora M, Öpik M. 2021. User-friendly bioinformatics pipeline gDAT (graphical downstream analysis tool) for analysing rDNA sequences. *Molecular Ecology Resources*, 21(4):1380–1392.
- Venter JC. 2010. Multiple personal genomes await. *Nature*, 464(7289):676–677.

- Verma M, Kulshrestha S, Puri A. 2017. Genome Sequencing. In: Keith JM, Hrsg. Bioinformatics: Volume I: Data, Sequence Analysis, and Evolution. 3–33, New York, NY: Springer DOI: 10.1007/978-1-4939-6622-6\_1.
- Vester BM, Dalsing BL, Middelbos IS, Apanavicius CJ, Lubbs DC, Swanson KS. 2009. Faecal microbial populations of growing kittens fed high- or moderate-protein diets. *Archives of Animal Nutrition*, 63(3):254–265.
- Vincent AT, Derome N, Boyle B, Culley AI, Charette SJ. 2017. Next-generation sequencing (NGS) in the microbiological world: How to make the most of your money. *Journal of Microbiological Methods*, 138:60–71.
- Wah Tang P, San Chua P, Kee Chong S, Saberi Mohamad M, Wen Choon Y, Deris S, Omatu S, Manuel Corchado J, Howe Chan W, Abdul Rahim R. 2015. A Review of Gene Knockout Strategies for Microbial Cells. *Recent Patents on Biotechnology*, 9(3):176–197.
- Walker AW, Lawley TD. 2013. Therapeutic modulation of intestinal dysbiosis. *Pharmacological Research*, 69(1):75–86.
- Wolters M, Ahrens J, Romaní-Pérez M, Watkins C, Sanz Y, Benítez-Páez A, Stanton C, Günther K. 2019. Dietary fat, the gut microbiota, and metabolic health - A systematic review conducted within the MyNewGut project. *Clinical Nutrition (Edinburgh, Scotland)*, 38(6):2504–2520.
- Wong HC, Wang P. 2004. Induction of viable but nonculturable state in *Vibrio parahaemolyticus* and its susceptibility to environmental stresses. *Journal of Applied Microbiology*, 96(2):359–366.
- Zakrzewski M, Proietti C, Ellis JJ, Hasan S, Brion M-J, Berger B, Krause L. 2017. Calypso: a user-friendly web-server for mining and visualizing microbiome–environment interactions. *Bioinformatics*, 33(5):782–783.
- Zhang J, Chiodini R, Badr A, Zhang G. 2011. The impact of next-generation sequencing on genomics. *Journal of genetics and genomics = Yi chuan xue bao*, 38(3):95–109.
- Zhao X, Zhang Z, Hu B, Huang W, Yuan C, Zou L. 2018. Response of Gut Microbiota to Metabolite Changes Induced by Endurance Exercise. *Frontiers in Microbiology*, 9:765.
- Zhou H, Yuan Y, Wang H, Xiang W, Li shenjie, Zheng H, Wen Y, Ming Y, Chen L, Zhou J. 2021. Gut Microbiota: A Potential Target for Cancer Interventions. *Cancer Management and Research*, 13:8281–8296.
- Zhu H, Zhang H, Xu Y, Laššáková S, Korabečná M, Neužil P. 2020. PCR past, present and future. *BioTechniques*, 69(4):317–325.
- Zhu Q, Jin Z, Wu W, Gao R, Guo B, Gao Z, Yang Y, Qin H. 2014. Analysis of the intestinal lumen microbiota in an animal model of colorectal cancer. *PloS One*, 9(6):e90849.
- Zwirzitz B, Wetzels SU, Dixon ED, Stessl B, Zaiser A, Rabanser I, Thalgueter S, Pinior B, Roch F-F, Strachan C, et al. 2020. The sources and transmission routes of microbial populations throughout a meat processing facility. *NPJ biofilms and microbiomes*, 6(1):26.

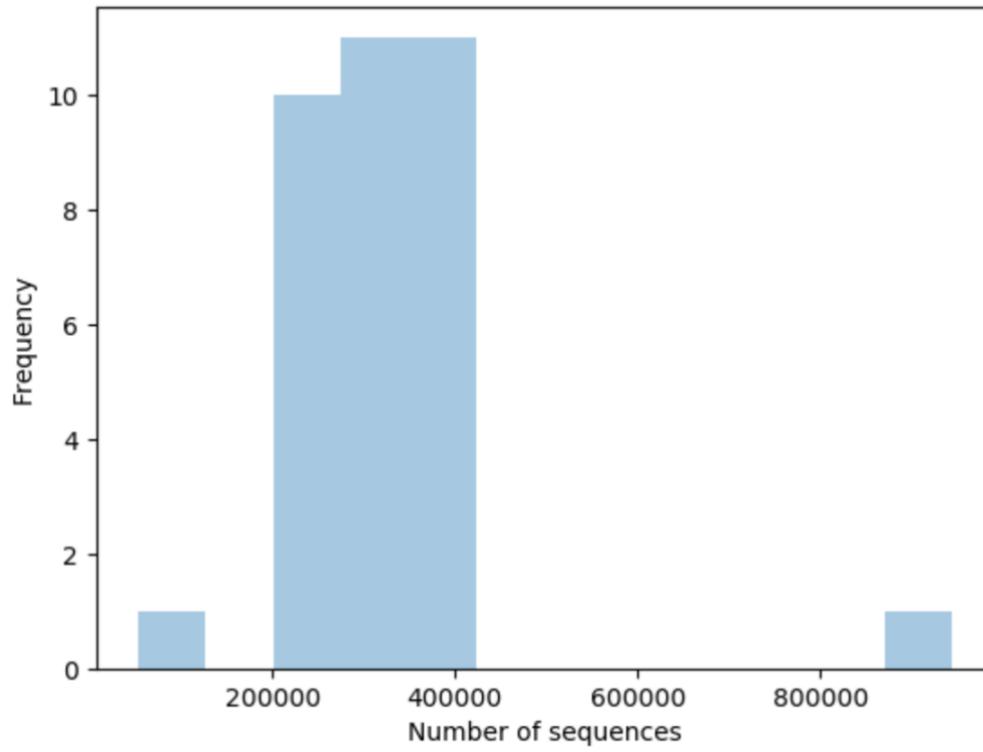
## 9. Abbildungs- und Tabellenverzeichnis

<b>Abb. 1</b> Systematische, taxonomische Einteilung der Bakterien .....	2
<b>Abb. 2</b> Sanger Sequenzierung .....	5
<b>Abb. 3</b> Übersicht über die Next-Generation-Sequenzierung .....	6
<b>Abb. 4</b> QIIME2 Viewer, Qualitätsbestimmung der Vorwärts- und Rückwärtssequenzen und visuelle Legung des Cutoffs .....	17
<b>Abb. 5</b> QIIME2 Viewer, der Ergebnisse der ASV Feature-Table.....	19
<b>Abb. 6</b> ASV Feature-ID Tabelle mit den zugehörigen Sequenzen .....	20
<b>Abb. 7</b> Qiime Viewer mit Werten des Shannon Parameters.....	22
<b>Abb. 8</b> ASV-ID und die zugehörigen taxonomische Klassifizierung.....	23
<b>Abb. 9</b> Taxa-Bar-Plot auf Phylum Ebene .....	30
<b>Abb. 10</b> Modifizierter Taxa-Bar-Plot auf Phylum Ebene .....	30
<b>Abb. 11</b> 3-D Emperor Abbildung der unweighted UniFrac-Distance .....	31
<b>Abb. 12</b> Boxplots für den Shannon Parameter .....	32
<b>Abb. 13</b> Boxplots für den Evenness Parameter .....	33
<b>Abb. 14</b> Relative Häufigkeit dargestellt auf Phylum Ebene .....	34
<b>Tab. 1</b> Vergleich der verschiedenen Illumina Sequenziermaschinen .....	8
<b>Tab. 2</b> Zusammenfassung des Metadatenfiles .....	18
<b>Tab. 3</b> Probenübersicht und die Zuordnung in die Lymphoma- oder gesunde Gruppe.....	27
<b>Tab. 4</b> Zusammenfassung der Sequenzanzahl.....	28
<b>Tab. 5</b> Überblick über die Alpha-Diversitäts Parameter .....	29
<b>Tab. 6</b> Die 30 häufigsten ASVs der Lymphoma- und der gesunden Gruppe .....	35



<b>Sample-ID</b>	<b>Sequence count</b>
S21L	945195
S20L	404528
S17L	396874
S28H	395669
S25H	391139
S18L	390200
S24H	389522
S01L	387946
S13L	381422
S26H	373976
S29H	370525
S04L	362933
S16L	346310
S06L	341926
S12L	340012
S08L	324987
S33H	322910
S32H	305343
S14L	297908
S27H	292135
S07L	290639
S15L	290186
S11L	276551
S23L	267619
S22L	264631
S30H	246687
S19L	235823
S03L	233320
S31H	227929
S05L	226643
S02L	216970
S10L	215688
S09L	200625
S34LK	51506
<b>Durchschnitt</b>	<b>323714</b>

**Supplement Tab. 3** Übersicht über den Probenamen (L= Lymphome und H= Healthy/Gesund), die Anzahl der Sequenzen pro Probe und den Mittelwert der Sequenzen pro Probe



**Supplement Abb. 1** Graphische Übersicht über die Anzahl der Sequenzen pro Probe

<b>Sample-ID</b>	<b>Barcode-Sequence</b>	<b>Sample Type</b>	<b>Gruppe</b>	<b>Behandlung</b>
<b>q2:types</b>	<b>categorical</b>	<b>categorical</b>	<b>categorical</b>	<b>categorical</b>
S01L		Sample	Lymphoma	no
S02L		Sample	Lymphoma	no
S03L		Sample	Lymphoma	treated
S04L		Sample	Lymphoma	no
S05L		Sample	Lymphoma	no
S06L		Sample	Lymphoma	no
S07L		Sample	Lymphoma	blood
S08L		Sample	Lymphoma	no
S09L		Sample	Lymphoma	no
S10L		Sample	Lymphoma	no
S11L		Sample	Lymphoma	no
S12L		Sample	Lymphoma	no
S13L		Sample	Lymphoma	no
S14L		Sample	Lymphoma	treated
S15L		Sample	Lymphoma	treated
S16L		Sample	Lymphoma	no
S17L		Sample	Lymphoma	no
S18L		Sample	Lymphoma	no
S19L		Sample	Lymphoma	no
S20L		Sample	Lymphoma	no
S21L		Sample	Lymphoma	no
S22L		Sample	Lymphoma	no
S23L		Sample	Lymphoma	no
S24H		Sample	Healthy	no_healthy
S25H		Sample	Healthy	no_healthy
S26H		Sample	Healthy	no_healthy
S27H		Sample	Healthy	no_healthy
S28H		Sample	Healthy	no_healthy
S29H		Sample	Healthy	no_healthy
S30H		Sample	Healthy	no_healthy
S31H		Sample	Healthy	no_healthy
S32H		Sample	Healthy	no_healthy
S33H		Sample	Healthy	no_healthy
S34LK		LK	LK	

**Supplement Tab. 4** Metadatenfile, bestehend aus Proben-ID, Barcode-Sequence, Sample Type, Gruppe und Behandlung.

Sample-ID	Rohsequenzen	Gefilterte Sequenzen	Gefilterte Sequenzen (%)	Kontrollierte Sequenzen	Kontrollierte Sequenzen (%)
S01L	387946	152294	39,26%	235652	60,74%
S02L	216970	57583	26,54%	159387	73,46%
S03L	233320	76724	32,88%	156596	67,12%
S04L	362933	141529	39,00%	221404	61,00%
S05L	226643	72353	31,92%	154290	68,08%
S06L	341926	123493	36,12%	218433	63,88%
S07L	290639	101957	35,08%	188682	64,92%
S08L	324987	115580	35,56%	209407	64,44%
S09L	200625	63819	31,81%	136806	68,19%
S10L	215688	83426	38,68%	132262	61,32%
S11L	276551	73819	26,69%	202732	73,31%
S12L	340012	121193	35,64%	218819	64,36%
S13L	381422	123924	32,49%	257498	67,51%
S14L	297908	99470	33,39%	198438	66,61%
S15L	290186	97661	33,65%	192525	66,35%
S16L	346310	104007	30,03%	242303	69,97%
S17L	396874	147020	37,04%	249854	62,96%
S18L	390200	135448	34,71%	254752	65,29%
S19L	235823	71350	30,26%	164473	69,74%
S20L	404528	142444	35,21%	262084	64,79%
S21L	945195	321721	34,04%	623474	65,96%
S22L	264631	76342	28,85%	188289	71,15%
S23L	267619	70089	26,19%	197530	73,81%
S24H	389522	157938	40,55%	231584	59,45%
S25H	391139	140292	35,87%	250847	64,13%
S26H	373976	137145	36,67%	236831	63,33%
S27H	292135	117365	40,17%	174770	59,83%
S28H	395669	153137	38,70%	242532	61,30%
S29H	370525	159155	42,95%	211370	57,05%
S30H	246687	98285	39,84%	148402	60,16%
S31H	227929	91648	40,21%	136281	59,79%
S32H	305343	117454	38,47%	187889	61,53%
S33H	322910	127292	39,42%	195618	60,58%
S34LK	51506	11047	21,45%	40459	78,55%
<b>Summe:</b>	11006277	3884004	34,69%	7122273	65,31%

**Supplement Tab. 5** Zusammenfassung der Sequenzanalyse von Roh- und qualitätskontrollierten Sequenzen

<b>Gesamt</b>	<b>Gesund</b>	<b>Lymphoma</b>
<i>Firmicutes</i> (47,99 %)	<i>Firmicutes</i> (60,44 %)	<i>Firmicutes</i> (42,57 %)
<i>Bacteroidota</i> (28,01 %)	<i>Bacteroidota</i> (22,41 %)	<i>Bacteroidota</i> (30,44 %)
<i>Fusobacteria</i> (10,45 %)	<i>Fusobacteria</i> (1,85 %)	<i>Fusobacteria</i> (14,19 %)
<i>Proteobacteria</i> (6,90 %)	<i>Proteobacteria</i> (1,89 %)	<i>Proteobacteria</i> (9,08 %)
<i>Actinobacteria</i> (5,69 %)	<i>Actinobacteria</i> (12,36 %)	<i>Actinobacteria</i> (2,79 %)
SUMME: 99,04 %	SUMME: 98,94 %	SUMME: 99,08 %

**Supplement Tab. 6** Die fünf dominanten Phyla im gesamten Probensatz mit der jeweiligen Summe und der direkte Vergleich von den Phyla der Lymphoma Gruppe und den der gesunden Gruppe.

<b>Genus</b>	<b>Gesamt</b>	<b>Gesund</b>	<b>Lymphoma</b>
<i>Bacteroidaceae</i>	17,97%	12,75%	20,24%
<i>Peptostreptococcaceae</i>	14,99%	17,00%	14,11%
<i>Lachnospiraceae</i>	11,42%	17,51%	8,78%
<i>Fusobacteriaceae</i>	10,44%	1,85%	14,17%
<i>Ruminococcaceae</i>	6,60%	9,75%	5,22%
<i>Coriobacteriaceae</i>	4,78%	10,15%	2,45%
<i>Prevotellaceae</i>	4,77%	5,13%	4,61%
<i>Clostridiaceae 1</i>	4,64%	2,97%	5,36%
<i>Burkholderiaceae</i>	3,61%	0,91%	4,78%
<i>Veillonellaceae</i>	2,45%	3,86%	1,84%
<i>Enterobacteriaceae</i>	2,41%	0,40%	3,29%
<i>Erysipelotrichaceae</i>	2,41%	4,40%	1,55%
<i>Tannerellaceae</i>	2,12%	2,91%	1,78%
<i>Streptococcaceae</i>	1,84%	0,58%	2,39%
<i>Marinifilaceae</i>	1,27%	0,25%	1,71%
<i>Rikenellaceae</i>	1,01%	0,80%	1,10%
<i>Enterococcaceae</i>	0,96%	0,13%	1,32%
<i>Family XIII</i>	0,92%	1,19%	0,80%
<i>Peptococcaceae</i>	0,83%	1,90%	0,36%
<i>Acidaminococcaceae</i>	0,70%	0,88%	0,62%
<i>Desulfovibrionaceae</i>	0,49%	0,44%	0,51%
<i>Barnesiellaceae</i>	0,45%	0,32%	0,51%
<i>Bifidobacteriaceae</i>	0,42%	1,27%	0,05%
<i>Helicobacteraceae</i>	0,35%	0,92%	0,11%

**Supplement Tab. 7** Die 24 häufigsten Familien des Probensatzes und deren Häufigkeit insgesamt, in der Lymphoma Gruppe und in der gesunden Gruppe

ASV-ID	Feature-ID	Dominum	Phylum	Classis	Ordo	Familia	Genus	Spezies
1	d059cf6b6634be17f15a238135743a8e	D_0_Bacteria	D_1_Firmicutes	D_2_Clostridia	D_3_Clostridiales	D_4_Peptostreptococcaceae	D_5_Peptoclostridium	D_6_uncultured bacterium
2	47959156085c7e6b7bdb022c16c749f5	D_0_Bacteria	D_1_Fusobacteria	D_2_Fusobacteriia	D_3_Fusobacteriales	D_4_Fusobacteriaceae	D_5_Fusobacterium	D_6_gut metagenome
3	e2985ddaea1c89944d7081554b1b0ef1	D_0_Bacteria	D_1_Firmicutes	D_2_Clostridia	D_3_Clostridiales	D_4_Clostridiaceae 1	D_5_Clostridium sensu stricto 1	
4	0f2f963e24b81d523800c9664cb8fc21	D_0_Bacteria	D_1_Bacteroidota	D_2_Bacteroidia	D_3_Bacteroidales	D_4_Bacteroidaceae	D_5_Bacteroides	
5	8cbd6dabb13df66486da45568f260de6	D_0_Bacteria	D_1_Firmicutes	D_2_Clostridia	D_3_Clostridiales	D_4_Peptostreptococcaceae	D_5_Paeniclostridium	
6	562f8e5fe4e773855042ddd638cdf26	D_0_Bacteria	D_1_Bacteroidota	D_2_Bacteroidia	D_3_Bacteroidales	D_4_Bacteroidaceae	D_5_Bacteroides	
7	e9f922b0323f3833ef6e7dae13f71d54	D_0_Bacteria	D_1_Bacteroidota	D_2_Bacteroidia	D_3_Bacteroidales	D_4_Bacteroidaceae	D_5_Bacteroides	D_6_Bacteroides stercoris ATCC 43183
8	a498e9b5a610147b50de9176cee83849	D_0_Bacteria	D_1_Actinobacteria	D_2_Coriobacteriia	D_3_Coriobacteriales	D_4_Coriobacteriaceae	D_5_Collinsella	
9	5397cf19d8e23d4b7ea55e91901f6899	D_0_Bacteria	D_1_Proteobacteria	D_2_Gammaproteobacteria	D_3_Enterobacteriales	D_4_Enterobacteriaceae	D_5_Escherichia-Shigella	
10	e18b13c0a48737c7e739e1f881c2c4b2	D_0_Bacteria	D_1_Bacteroidota	D_2_Bacteroidia	D_3_Bacteroidales	D_4_Bacteroidaceae	D_5_Bacteroides	
11	ba42c01bc3eb875307a7339eb249d9ca	D_0_Bacteria	D_1_Fusobacteria	D_2_Fusobacteriia	D_3_Fusobacteriales	D_4_Fusobacteriaceae	D_5_Fusobacterium	
12	10a52f96ba29e1ee2c6893459e56a7b2	D_0_Bacteria	D_1_Bacteroidota	D_2_Bacteroidia	D_3_Bacteroidales	D_4_Bacteroidaceae	D_5_Bacteroides	
13	4a32ec31f42357144d7cf0ab6956aea5	D_0_Bacteria	D_1_Bacteroidota	D_2_Bacteroidia	D_3_Bacteroidales	D_4_Tannerellaceae	D_5_Parabacteroides	D_6_Parabacteroides merdae
14	848eca9cd5555c4edf12439ecf66475f	D_0_Bacteria	D_1_Actinobacteria	D_2_Coriobacteriia	D_3_Coriobacteriales	D_4_Coriobacteriaceae	D_5_Collinsella	D_6_Collinsella tanakaei YIT 12063
15	9c878416a5d9ff309ee397443f60005b	D_0_Bacteria	D_1_Firmicutes	D_2_Bacilli	D_3_Lactobacillales	D_4_Streptococcaceae	D_5_Streptococcus	
16	6ace990f3b5cd6144c0d537a32b59898	D_0_Bacteria	D_1_Firmicutes	D_2_Clostridia	D_3_Clostridiales	D_4_Peptostreptococcaceae	D_5_Peptostreptococcus	
17	17a7a5de2b84fd511f678eef9b283548	D_0_Bacteria	D_1_Proteobacteria	D_2_Gammaproteobacteria	D_3_Betaproteobacteriales	D_4_Burkholderiaceae	D_5_Sutterella	D_6_uncultured bacterium
18	2605da6501fc62aaec0572060a1e32f2	D_0_Bacteria	D_1_Proteobacteria	D_2_Gammaproteobacteria	D_3_Pseudomonadales	D_4_Pseudomonadaceae	D_5_Pseudomonas	
19	9cb0009ac96b0463ed88ee12266ed80b	D_0_Bacteria	D_1_Firmicutes	D_2_Erysipelotrichia	D_3_Erysipelotrichales	D_4_Erysipelotrichaceae	D_5_Holdemanella	
20	651c539c68bd57db947e0bd8376729da	D_0_Bacteria	D_1_Firmicutes	D_2_Clostridia	D_3_Clostridiales	D_4_Lachnospiraceae	D_5_Blautia	
21	b8a055ec492259c41ac7bc32982025ef	D_0_Bacteria	D_1_Firmicutes	D_2_Clostridia	D_3_Clostridiales	D_4_Ruminococcaceae	D_5_Negativibacillus	

22	cfa10b18df13dbab9 a47bb7cffffa655	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Ruminococcaceae	D_5	Ruminococcaceae UCG-013	D_6	uncultured bacterium
23	3419864bcd1ba4ffc 7333345c5259c81	D_0	Bacteria	D_1	Proteobacteria	D_2	Gammaproteobacteria	D_3	Betaproteobacteriales	D_4	Burkholderiaceae	D_5	Sutterella	D_6	uncultured bacterium
24	6ce6756e4431b100 51deb413c588ee68	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Lachnospiraceae	D_5	Blautia		
25	c6f94183db3eccce 2f9a9618d63142c4	D_0	Bacteria	D_1	Proteobacteria	D_2	Gammaproteobacteria	D_3	Betaproteobacteriales	D_4	Burkholderiaceae	D_5	Parasutterella	D_6	gut metagenome
26	3cf8653b4ea2feb00 7ec583dddc082f4	D_0	Bacteria	D_1	Bacteroidota	D_2	Bacteroidia	D_3	Bacteroidales	D_4	Bacteroidaceae	D_5	Bacteroides		
27	8dff08e5ca4689082 5448c7622d8288a	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Lachnospiraceae				
28	54d8ed8275f89df38 f123e98310d1ac3	D_0	Bacteria	D_1	Fusobacteria	D_2	Fusobacteriia	D_3	Fusobacteriales	D_4	Fusobacteriaceae	D_5	Fusobacterium		
29	dec18fb177e0b0d0 aaf24e416374e8bf	D_0	Bacteria	D_1	Firmicutes	D_2	Bacilli	D_3	Lactobacillales	D_4	Enterococcaceae	D_5	Enterococcus		
30	b0d94dbbde5bee68 3dbdae3836438c62	D_0	Bacteria	D_1	Bacteroidota	D_2	Bacteroidia	D_3	Bacteroidales	D_4	Marinifilaceae	D_5	Odoribacter	D_6	uncultured bacterium
31	4c32ac1bbc24f5fb6 f380869955b43af	D_0	Bacteria	D_1	Bacteroidota	D_2	Bacteroidia	D_3	Bacteroidales	D_4	Rikenellaceae	D_5	Alistipes		
32	6cdedf031fe142306 f65ed664b73e085	D_0	Bacteria	D_1	Bacteroidota	D_2	Bacteroidia	D_3	Bacteroidales	D_4	Bacteroidaceae	D_5	Bacteroides		
33	081d30dbfba5bc7e 48cae0750218a9ce	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Family XIII	D_5	[Eubacterium] brachy group	D_6	uncultured bacterium
34	f6540507f6678f6dfe 995252fcd55247	D_0	Bacteria	D_1	Bacteroidota	D_2	Bacteroidia	D_3	Bacteroidales	D_4	Prevotellaceae	D_5	Alloprevotella		
35	each342ef5c436d7f fc783492911fec1	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Lachnospiraceae	D_5	Blautia		
36	15b294fde907740e 76bce0df267bdcab	D_0	Bacteria	D_1	Firmicutes	D_2	Negativicutes	D_3	Selenomonadales	D_4	Acidaminococcaceae	D_5	Phascolarctobacterium		
37	0326c90f8184dbaa cf8c6bed95bb6a49	D_0	Bacteria	D_1	Bacteroidota	D_2	Bacteroidia	D_3	Bacteroidales	D_4	Prevotellaceae	D_5	Prevotella 9		
38	875a772e09e60ae4 f0a157eedaaa8699	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Peptococcaceae	D_5	Peptococcus	D_6	uncultured bacterium
39	e416dbe5c5986529 aa9e2d2858872905	D_0	Bacteria	D_1	Firmicutes	D_2	Negativicutes	D_3	Selenomonadales	D_4	Veillonellaceae	D_5	Megasphaera		
40	bc28ef338ab1fd58c 266355a65b31b1b	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Lachnospiraceae	D_5	Blautia		
41	a12eebe02d259cc7 0bcf3367feb7663f	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Ruminococcaceae	D_5	Faecalibacterium		
42	8fec5cd6de304ce3 26d10cfb05e75c94	D_0	Bacteria	D_1	Actinobacteria	D_2	Coriobacteriia	D_3	Coriobacteriales	D_4	Coriobacteriaceae	D_5	Collinsella		
43	40df354e940d0754 37097e7caf4fd4dbb	D_0	Bacteria	D_1	Firmicutes	D_2	Negativicutes	D_3	Selenomonadales	D_4	Veillonellaceae	D_5	Megamonas	D_6	uncultured bacterium
44	f8236729695fbc393 f54e896b20dbde0	D_0	Bacteria	D_1	Bacteroidota	D_2	Bacteroidia	D_3	Bacteroidales	D_4	Bacteroidaceae	D_5	Bacteroides	D_6	Bacteroides massiliensis B84634 = Timone 84634 =

45	a4ed5f147748e0f5a be111eda38eb264 4b75e344fb64e733	D_0	Bacteria	D_1	Firmicutes	D_2	Bacilli	D_3	Lactobacillales	D_4	Streptococcaceae	D_5	Streptococcus	
46	118250199fe8b285 50edfa82ed1a59b3 74f6143f08751b7a	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Lachnospiraceae			
47	20945b5bae49c952 529af09bf4139aee	D_0	Bacteria	D_1	Proteobacteria	D_2	Alphaproteobacteria	D_3	Rickettsiales	D_4	Anaplasmataceae	D_5	Wolbachia	
48	5c49ee50a755f26c 3b6c83be787d4268	D_0	Bacteria	D_1	Firmicutes	D_2	Erysipelotrichia	D_3	Erysipelotrichales	D_4	Erysipelotrichaceae	D_5	Catenibacterium	
49	5c1450274bf34e1a a04e478e2b708464	D_0	Bacteria	D_1	Proteobacteria	D_2	Gamma proteobacteria	D_3	Enterobacteriales	D_4	Enterobacteriaceae	D_5	Escherichia-Shigella	
50	fc12394b7398533a e13cc3109f81443d	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Lachnospiraceae	D_5	[Ruminococcus] torques group	
51	fe2a923646e26a4d e42be44924f47313 c962eef9b7a747f2f	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Ruminococcaceae	D_5	Ruminococcaceae UCG-005	
52	a89aebd9933711c 3475f9cf6d2fedf13b	D_0	Bacteria	D_1	Bacteroidota	D_2	Bacteroidia	D_3	Bacteroidales	D_4	Bacteroidaceae	D_5	Bacteroides	
53	e239ad22073886 c9963c8ad583ab6c 1dc4e4cb59d21d25	D_0	Bacteria	D_1	Bacteroidota	D_2	Bacteroidia	D_3	Bacteroidales	D_4	Prevotellaceae	D_5	Prevotella 9	
54	fff8d4c13e9ea600d 9916594d8cc86b7	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Peptostreptococcaceae	D_5	Clostridioides	
55	710a48984abcd524 a92136313ecf1e42	D_0	Bacteria	D_1	Bacteroidota	D_2	Bacteroidia	D_3	Bacteroidales	D_4	Prevotellaceae	D_5	Prevotella 9	D_6_Megasphaera elsdenii
56	9aec9299b9d0eae 9d6d64766bebeab1	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Ruminococcaceae	D_5	Negativibacillus	D_6_uncultured bacterium
57	273a8add2b7469d7 bfea55fc1a4692e9	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Lachnospiraceae	D_5	Blautia	
58	6fa3d5a942b6b79f7 cddd244f685389	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Lachnospiraceae	D_5	Lachnoclostridium	D_6_Lachnospiraceae bacterium 2_1_46FAA
59	0b38e58cd6164ace 59b5781e4085d2be abbe46051e219793	D_0	Bacteria	D_1	Bacteroidota	D_2	Bacteroidia	D_3	Bacteroidales	D_4	Bacteroidaceae	D_5	Bacteroides	D_6_Bacteroides massiliensis B84634 = Timone 84634 = DSM 17679 = JCM 13223
60	340870f1cc8f8ea7 601bd0d8d2edb3e8 fc7de94e9454e23a	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Ruminococcaceae	D_5	Fournierella	
61	35d64d33280398d4 870007ceb6b9af36	D_0	Bacteria	D_1	Bacteroidota	D_2	Bacteroidia	D_3	Bacteroidales	D_4	Bacteroidaceae	D_5	Bacteroides	
62	90c4742ba6ba926e 00ffec3690effeaa	D_0	Bacteria	D_1	Bacteroidota	D_2	Bacteroidia	D_3	Bacteroidales	D_4	Helicobacteraceae	D_5	Helicobacter	
63	870007ceb6b9af36 90c4742ba6ba926e 00ffec3690effeaa	D_0	Bacteria	D_1	Epsilonbacterae ota	D_2	Campylobacteria	D_3	Campylobacteriales	D_4	Helicobacteraceae	D_5	Helicobacter	D_6_uncultured organism
64	870007ceb6b9af36 90c4742ba6ba926e 00ffec3690effeaa	D_0	Bacteria	D_1	Firmicutes	D_2	Negativicutes	D_3	Selenomonadales	D_4	Veillonellaceae	D_5	Megamonas	
65	870007ceb6b9af36 90c4742ba6ba926e 00ffec3690effeaa	D_0	Bacteria	D_1	Bacteroidota	D_2	Bacteroidia	D_3	Bacteroidales	D_4	Prevotellaceae	D_5	Prevotella 9	

66	383109f17f4bc06c3 2a02adf1e62c130	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Lachnospiraceae	D_5	Blautia	
67	1175f6e5cd8bd6fd2 45874b741a8706b	D_0	Bacteria	D_1	Tenericutes	D_2	Mollicutes	D_3	Anaeroplasmatales	D_4	Anaeroplasmataceae	D_5	Anaeroplasma	D_6_gut metagenome
68	e38ea3aa742603a9 c681e9c8ccb4d497	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Lachnospiraceae	D_5	Blautia	
69	a46701ffeb5da6a4c ea3b03c7c92004c	D_0	Bacteria	D_1	Bacteroidota	D_2	Bacteroidia	D_3	Bacteroidales	D_4	Prevotellaceae	D_5	Prevotella 9	
70	0a706046ee1e387c fd4fca65d77ccf0f	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Lachnospiraceae	D_5	Blautia	
71	5bcfb9e80dbf8d605 565de1b3b995a55	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Lachnospiraceae	D_5	[Ruminococcus] torques group	
72	55f7278323534ee9 8161daad01bae541	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Lachnospiraceae	D_5	Blautia	
73	224ce5303a4d3a67 d1bcefc3aec53fac	D_0	Bacteria	D_1	Fusobacteria	D_2	Fusobacteriia	D_3	Fusobacteriales	D_4	Fusobacteriaceae	D_5	Fusobacterium	
74	d4c59d12129cc0a0 a9ec27ad97d9a6c6	D_0	Bacteria	D_1	Firmicutes	D_2	Clostridia	D_3	Clostridiales	D_4	Lachnospiraceae			
75	4e38b9191da004ae 6ff37a0df967440e	D_0	Bacteria	D_1	Actinobacteria	D_2	Coriobacteriia	D_3	Coriobacteriales	D_4	Coriobacteriaceae	D_5	Collinsella	
76	8c881dfb0ba5d4bb c64fca1427bee95b	D_0	Bacteria	D_1	Proteobacteria	D_2	Deltaproteobacteria	D_3	Desulfovibrionales	D_4	Desulfovibrionaceae	D_5	Desulfovibrio	D_6_gut metagenome
77	0d96aeec02003e0 dd18739836e392e6	D_0	Bacteria	D_1	Bacteroidota	D_2	Bacteroidia	D_3	Bacteroidales	D_4	Bacteroidaceae	D_5	Bacteroides	D_6_Bacteroides pyogenes F0041
78	11e1d53e750c9304 ac54a1ace36e1817	D_0	Bacteria	D_1	Proteobacteria	D_2	Gammaproteobacteria	D_3	Enterobacteriales	D_4	Enterobacteriaceae	D_5	Serratia	
79	c7a7a3071f89b194 7f7a7c8ff8a72c7f	D_0	Bacteria	D_1	Bacteroidota	D_2	Bacteroidia	D_3	Bacteroidales	D_4	Bacteroidaceae	D_5	Bacteroides	D_6_Bacteroides stercoris ATCC 43183
80	2da9146e7091662b 98533fffc136a9a0	D_0	Bacteria	D_1	Actinobacteria	D_2	Actinobacteria	D_3	Bifidobacteriales	D_4	Bifidobacteriaceae	D_5	Bifidobacterium	

**Supplement Tab. 8** Die 80 häufigsten ASVs des gesamten Probensatzes, sortiert nach absteigender Häufigkeit, mit Feature-ID und Taxonomie von Domäne (D\_0) bis Spezies (D\_6)

