

# Bericht zum Projekt

## **Evaluierung von NGS zur molekularen Epidemiologie bakterieller Pathogene**

### **Inhalt**

1	Einleitung und Projektanforderungen .....	2
2	Evaluierung verschiedener Methoden .....	6
2.1	Etablierung von Methoden zur Herstellung geeigneter DNA.....	6
2.1.1	DNA Aufreinigungsverfahren.....	7
3	Vergleich unterschiedlicher Sequenziermethoden.....	15
3.1	Illumina Sequenzierung.....	15
3.2	Pacific Bioscience.....	15
3.3	Nanopore Sequenzierung und Etablierung am LGL.....	16
3.3.1	Sequenzierläufe .....	19
3.3.2	Sequenzenauswertung.....	24
4	Zusammenfassung und Schlussfolgerungen.....	31
5	Literatur .....	32

## 1 Einleitung und Projektanforderungen

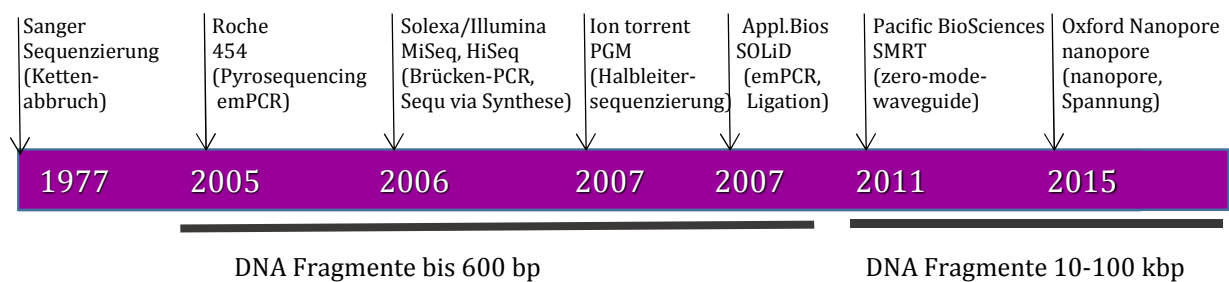
Multi Locus Sequence Typing (MLST) und Sequenziermethoden der nächsten Generation (Next Generation Sequencing; NGS) gehören zu den entscheidenden Methoden in der molekulare Epidemiologie. NGS Technologien werden nicht nur in der Wissenschaft genutzt, sondern haben auch Einzug in medizinische Einrichtungen gehalten und werden mehr und mehr für diagnostische Zwecke und im öffentlichen Gesundheitsdienst eingesetzt. Insbesondere bei der Aufklärung von Ausbrüchen infektiöser Pathogene, entweder im human- oder veterinärmedizinischen Bereich und auch um lebensmittelbedingte Infektionsketten abzuklären.

Ziel des Projektes war die Etablierung und Ausarbeitung von standardisierten, qualitätsgesicherten Abläufen (workflows) für Infektionserreger am LGL. Hierzu gehört zunächst der Vergleich von Methoden zur Aufreinigung von DNA, da die Quantität und Qualität bei den Methoden zur Sequenzierung langer Reads eine große Rolle spielen. Ein Schwerpunkt des Projektes lag auf der Etablierung von Oxford Nanopore Technologie (ONT), um Erkenntnisse darüber zu erhalten, welche der Technologien, die lange Sequenzreads liefern, am besten für die Routinediagnostik geeignet sind. Während Pacific Bioscience single-molecule read-time (SMRT) Reads bis zu 20 kb liefert, können mit ONT Reads von bis zu 100 kb hergestellt werden. Vorteil der SMRT Methodik ist, dass Reads eine höhere Präzision aufweisen als ONT. Auf der anderen Seite sind für die SMRT Methodik aber wesentlich größere DNA Mengen erforderlich und die Kosten der Read-Generierung sind höher als für ONT.

Im Folgenden ist eine kurze Übersicht über die unterschiedlichen Sequenziermethoden gegeben, die später im Bericht noch weiter erläutert werden. In den letzten zwei Dekaden sind eine Reihe von Technologien im Bereich NGS entwickelt und kommerzialisiert worden (siehe Abb. 1), die die massive Parallelsequenzierung vieler DNA Fragmente erlauben und somit die Sequenzierung von DNA Fragmenten in einem Bruchteil der Zeit und kostengünstiger als die bislang übliche Sanger Sequenzierung bewerkstelligen können. Die verschiedenen Technologien nutzen unterschiedliche Librarypreparationen und unterschiedliche Signalerfassung, was zu Vor- oder Nachteilen führen kann (Heather and Chain 2016; Levy and Myers 2016). Bedingt durch die unterschiedliche Methodik in Probenpräparation und Signalerfassung sind Sequenzierfehler mehr oder weniger häufig, besonders im Bereich von

homopolymeren Regionen (Quail, Smith et al. 2012). Sequenziermethoden unterscheiden sich auch in der Länge der sequenzierten DNA Fragmente: während die Sequenziertechnologien Roche 454, Ion torrent, Solid und Illumina kurze Fragmente sequenzieren (bis zu 600 bp), können die zuletzt entwickelten Sequenziertechnologien wie Pacific BioScience SMRT Technologie und Nanopore Technologie lange (10-20kbp, SMRT) bis ultralange (bis 100 kbp, Nanopore) DNA Fragmente in einem Stück sequenzieren. Letzteres hat große Vorteile bei der *de novo* Zusammensetzung von Ganzgenomen.

Unter den Technologien, die zur Sequenzierung kurzer DNA Fragmente geeignet sind (Abb. 1), ist die Illumina Technologie ganz klar als Marktführer hervorgegangen (Reuter, Spacek et al. 2015) (Hodzic, Gurbeta et al. 2017). Dies ist auf die vermutlich auf die Präzision der erhaltenen Reads, Zuverlässigkeit der Brückenamplifikation, Einfachheit der Libraryherstellung und Kostengünstigkeit zurückzuführen.



**Abbildung 1:** Zeitverlauf der Entwicklung von Sequenziertechnologien

Im Folgenden ist eine kurze Darstellung der drei von uns eingesetzten Methoden gegeben:

*Illuminasequenzierung.* Für Illuminasequenzierung gibt es unterschiedliche Formen der Librarypreparation, die praktischste Art der Herstellung ist mit Nextera XT kits. Vorteile sind i) es werden sehr geringe Mengen an DNA benötigt (1 ng total DNA), ii) die DNA wird gleichzeitig fragmentiert und ein Adapter angefügt, der die komplementäre Sequenz eines PCR Primers enthält, iii) über einen kurzen PCR-Amplifizierungsschritt werden Indices und weitere Adapter angefügt, die es erlauben, die DNA Fragmente auf der Flowcell zu binden. Die Indexflexibilität erlaubt es, bis zu 96 Proben parallel zu sequenzieren. Die fertigen und normalisierten Libraries werden auf die Flowcell gegeben, wo die Brückenamplifikation mit Clusterbildung stattfindet. Die

Flowcell ist mit einem Rasen komplementärer Oligos bedeckt, an den die DNA Fragmente binden können. Es erfolgt eine Clusterbildung, indem jedes DNA Fragment über Brücken-PCR amplifiziert wird. Dies geschieht, damit die Signalerfassung ermöglicht wird. Der eigentliche Sequenzierungsschritt (Sequenzierung durch Synthese) läuft – je nach Größe der Flowcell – in 24 – 72 Stunden ab. Die sequenzierten Reads werden noch im Gerät demultiplexed, d.h. alle Reads mit identischem Index werden in einem Ordner zusammengefasst, und die Adaptersequenzen abgeschnitten.

Für Illuminasequenzierung stehen unterschiedliche große Geräten zur Verfügung, je nach Bedarf der Größe der zu sequenzierenden Organismen. Für bakterielle Genome eignet sich der MiSeq, eine benchtop Version, die am LGL seit 2013 zur Verfügung steht.

*Pacific Biosciences SMRT Technologie.* Bei der SMRT Technologie werden Einzelstrang-Hairpin-Adapter an die Enden von langkettigen DNA Fragmenten ligiert, wodurch seitlich begrenzte DNA Stränge entstehen, die sogenannten ‚SMRT-bells‘. Diese DNA Stränge werden in der sogenannten ‚zero-mode-waveguide‘ sequenziert, ebenfalls durch Sequenzierung-durch-Synthese. Die zero-mode-waveguide ist eine Nanozelle, in der ein einzelnes DNA Polymerase Molekül am Boden immobilisiert ist, und die gerade gross genug ist, um Lichtsignale erkennen zu lassen und zu messen. Dadurch, dass die DNA Fragmente als Ringmolekül vorliegen (SMRT-bell), kann jeder Strang mehrfach sequenziert werden. Dadurch können Fehler in der Basenfolge erkannt und ausgemerzt werden, was die Genauigkeit der Sequenzierung erhöht. Durch neue Methoden können neuerdings Proben mit einem Index versehen werden, d.h. mehrere Proben in einem Lauf sequenziert werden (Multiplexed), wodurch die Sequenzierkosten erheblich verringert wurden.

Pacific Bioscience Sequenziergeräte sind nicht als benchtop Modell zu haben, erfordern also erheblich mehr Platz und kosten erheblich mehr als Illumina Sequenzierer. Sie stehen meist in Sequenziereinheiten in Firmen oder Universitäten zur Verfügung.

*Oxford Nanopore Technologie (ONT).* Die Sequenzierung mit ONT beruht auf einem völlig anderen Prinzip. Für ONT wird eine Flowcell benutzt, die ca. 2,000 Nanoporen enthält. Eine Membrane (deren Beschaffenheit nicht bekannt ist) unterteilt die Nanoporen enthaltenden Zellen waagrecht und jede Zelle ist mit einer Einzelpore

ausgestattet. Die Basenfolge in der DNA wird mit Hilfe von Spannungsänderungen in der Nanopore ermittelt, die durch die Anwesenheit der DNA entsteht und je nach Basenabfolge unterschiedliche Schwankungen hervorruft. Abgesehen von DNA kann man mit ONT auch RNA direkt sequenzieren, was mit den anderen Technologien bisher nicht möglich ist.

Die DNA Vorbereitung erfolgt ebenfalls durch sogenannte Libraries; diese Librarypreparationen können mit oder ohne Fragmentierung der DNA oder mit oder ohne PCR Zwischenschritt erfolgen. Bei dieser Methode ist zu bedenken, dass sehr langkettige DNA schwierig zu verarbeiten ist, da es beim Pipettieren zu Scherkräften kommt, die sehr lange DNA Stücke fragmentieren kann. Es werden wie bei den anderen Methoden Adapter an die DNA ligiert, und es können auch Indices verwendet werden, um mehrere DNA Proben gleichzeitig zu sequenzieren. Während der Librarypreparation wird ein hydrophober Arm und ein Motorenzym an dem DNA Strang angebracht, die die DNA an die Pore binden und für den Durchschub der DNA durch die Pore sorgen. Gemessen werden bei dieser Technologie Spannungsschwankungen in der Pore, die durch Anwesenheit der DNA Basen verursacht werden, und die später in eine Basenabfolge übersetzt werden können.

Die oben gemachten Ausführungen zeigen, dass eine enorm wichtige Voraussetzung bei der Vorbereitung und Verarbeitung der Proben die Qualität und Quantität der DNA ist, welche besonders bei den Technologien, bei denen langkettige DNA Fragmente sequenziert werden sollen, eine ganz entscheidende Rolle spielt. Im Rahmen des Projektes wurden daher zunächst verschiedene DNA Aufreinigungsmethoden verglichen.

Ein weiterer Aspekt des Projektes beschäftigte sich mit Methoden der Sequenzauswertung und welche Software am besten geeignet ist. Während die Assemblierung von bakteriellen Kerngenomen auf Grund vergleichbarer Referenzsequenzen weniger eine Herausforderung darstellt, ist die Rekonstruktion der interessanten genomischen „Nicht-Kern“-Regionen (z.B. Plasmide oder andere genomische Regionen, die evtl. mobile Pathogenitätsfaktoren enthalten) häufig auf *de novo* Assemblierung angewiesen, da keine Referenzsequenz-geleiteten Strategien eingesetzt werden können. Es ist z.B. bekannt, dass die Aufnahme von Pathogenitätsfaktoren (z.B. Toxingene) durch Phagen oder andere mobile genomische Elemente

(Transposons) induziert werden kann (Jackson, Vinatzer et al. 2011; Moon, Park et al. 2016). Dies kann zu Veränderungen der Pathogenität von Infektionserregern führen. Es sind einige Softwares auf dem Markt, die eine verbesserte Assemblierung von Kern- und Nicht-Kern-Regionen in Genomen bakterieller Erreger versprechen. Dies wird z.T. durch Hybridassemblierung von lang- und kurzkettigen DNA Sequenzen gemacht. Proof-of-Principle Daten wurden bereits für Borrelien erhoben (Margos et al. 2017; Gofton et al. 2018). Es ist zu betonen, dass Borrelien sich aufgrund ihrer besonderen Genomstruktur besonders gut für solche Untersuchungen eignen. Zum einen sind sehr gute Referenzsequenzen für einige Isolate vorhanden, zum anderen stellt das aus Chromosom und Plasmiden bestehende Genom eine besondere Herausforderung für Assemblierungssoftware dar.

Im Rahmen des hier beschriebenen Projektes wurden diese neuen Techniken an bakteriellen Erregern getestet, im Besonderen an *Borrelia burgdorferi* sensu stricto und an Corynebakterien.

## **2 Evaluierung verschiedener Methoden**

Die folgenden Technologien sollten getestet und miteinander verglichen werden:

- Vergleich verschiedener DNA Aufreinigungsmethoden bezüglich Quantität und Qualität der DNA.
- Vergleichende Untersuchungen von kurz- und langkettigen Sequenziermethoden wie z.B. Illumina MiSeq, PacBioscience SMRT oder ONT an ausgewählten Proben.
- Austestung unterschiedlicher Assemblierungssoftware, die auch Hybridassembly von langen und kurzen Reads zulassen wie z.B. SPAdes, CLC Genomics Workbench, DNASTar genomics Module

### **2.1 Etablierung von Methoden zur Herstellung geeigneter DNA**

NGS stellt qualitative höhere Anforderungen an die Probenvorbereitung als konventionelle oder real-time PCR. Um optimale Protokolle für die DNA Aufreinigung zu erarbeiten, wurden entsprechende Protokolle ausgetestet. Am Nationalen

Referenzzentrum für Borrelien ist seit einigen Jahren eine Methode zur DNA Aufreinigung im Einsatz, die auf der Anwendung von Magnetpartikeln beruht. Das der Abteilung zur Verfügung stehende Gerät der Firma Promega hat sich als hervorragend geeignet für die quantitative DNA Aufreinigung aus Borrelien Kulturen herausgestellt. Während DNA Aufreinigung mittels Magnetpartikeln sehr reine DNA liefern kann, kann es jedoch bei der Aufreinigung zur Fragmentierung der DNA kommen, wobei Fragmente von einer Länge von 50-80 kb erhalten werden.

Seit kurzem ist eine neue DNA Extraktionstechnologie auf dem Markt, bei dem die DNA an eine Matrix in einer Pipettenspitze gebunden wird. Proteine und andere Moleküle werden herausgewaschen und die DNA anschließend extrahiert. Lt. Angaben des Herstellers soll diese Methode besonders schonend und effizient sein, hochaufgereinigte DNA, frei von Kontaminanten, liefern und die Extraktion von DNA Molekülen bis zu einer Länge von 400 kb erlauben.

Im Rahmen des Projektes wurden diese Methoden miteinander verglichen und die Ergebnisse werden im Folgenden dargestellt.

## 2.1.1 DNA Aufreinigungsverfahren

### 2.1.1.1 Säulen und Magnetobeads

Es wurde zunächst ein Vergleich der DNA Aufreinigung mittels Säulen (High Pure PCR Template Preparation Kit, Firma Roche) und mittels Magnetobeads im Maxwell Gerät (Maxwell® 16 LEV Blood DNA Kit, Firma Promega) gemacht. Es wurden 10 ml einer dicht bewachsenen Borrelienkulturen gemäß der Angaben des Herstellers gereinigt. Für die Aufreinigung mittels Maxwell (Promega) wurden zwei unterschiedliche Methoden verwendet: (1) Inkubation in CTAB Puffer; (2) Inkubation in Lyse Puffer (siehe Anhang). Diese beiden Methoden unterscheiden sich durch den benutzten Lysepuffer.

**Tabelle 1** Vergleich DNA Aufreinigung mittels Säulen oder Magnetobeads

Sample ID	Methode	ng/ul	A260	A280	260/280	260/230
PBi-1	Roche	14,23	0,285	0,201	1,42	0,71
25015-1	Roche	24,90	0,498	0,337	1,48	0,79
PBi-1	Maxwell 1	71,14	1,423	0,694	2,05	2,22
PBi-1	Maxwell 2	101,67	2,033	1,056	1,93	2,21

Die Konzentration der gereinigten DNA mittels Säulen war gering (14 – 25 ng/ul) und die Reinheit ließ zu wünschen übrig. Darüber hinaus führten Versuche, DNA aus Blut zu reinigen oft zu Problemen durch Verstopfen der Säulen.

Die photometrische Messung von Proteinen wird bei einer Wellenlänge von 260 nm, die von Nukleinsäuren bei 280 nm und die von Zuckern, Phenolen oder Salzen bei 230 nm gemessen. Die 260/280 Werte geben Auskunft über die Reinheit der gereinigten DNA bzw. die Anwesenheit von kontaminierenden Proteinen während die 260/230 Werte Auskunft über kontaminierende Salze, Kohlenhydrate oder andere Stoffe geben. 260/280 Werte von 1.8-2.0 und 260/230 Werte von  $\geq 2.0$  weisen auf sehr reine DNA hin. Wie aus der o.a. Tabelle ersichtlich ist, waren bei Säulenaufreinigung noch erheblich Kontaminanten in der DNA zu finden (260/280 1,4; 260/230 0,7), während die Aufreinigung über Magnetobeads erheblich bessere Werte aufwies; hier lag das Verhältnis von 260/280 bei den gewünschten Werten zwischen 1.8 – 2.0 sowie 260/230 Werte von  $>2$ .

DNA Aufreinigung mittels Magnetobeads brachte somit sehr viel bessere Ergebnisse. Mit dieser Methode ließen sich einerseits höhere DNA Konzentrationen (71-101 ng/ul) und auch eine wesentliche Verbesserung der Reinheit der DNA erzielen. Es wäre noch zu bemerken, dass die Methode 2 der Maxwell Aufreinigung besser abschnitt bei der Menge der gereinigten DNA, deshalb wurde im weiteren Verlauf des Projektes diese Methode weiterverwendet.

Kostenmäßig liegt die DNA Reinigung mittels Magnetobeads zwar wesentlich über dem der Säulenaufreinigung, diese Kosten sind durch die wesentlich besseren Ergebnisse aber vertretbar. Roche Kit Euro 196,-/100 Proben = Euro 1.96 pro Probe; Promega Kit Euro 240,-/48 Proben = Euro 5.10 pro Probe.



Für die zusätzliche Aufreinigung extrahierter genomischer DNA, d.h. wenn die 260/280 oder 260/230 Werte auf Verunreinigungen hinweisen, können entweder Ampurebeads (Agilent) oder auch Kits der Firmen Zymo oder Machery und Nagel verwendet werden. Ampurebeads sollen auch die Möglichkeit bieten, ‚kleine‘ DNA Fragmente (<2000 bp) aus Libraries zu entfernen (siehe: 3.3.1 Sequenzierläufe).

Die unterschiedlichen Methoden wurden im Rahmen des Projektes ausgetestet, entweder mit genomische *Borrlia burgdorferi* B31 DNA oder mit Lambda DNA. Die Reinigung der DNA wurde entweder genau nach Anweisung der Hersteller (Agilent, Zymo oder Machery und Nagel) oder wie in den entsprechenden Protokollen auf der Nanopore Website angegeben (Ampure Beads mit PEG Solution). Bei den Methoden mit Ampurebeads wurden sowohl im Überstand als auch im Eluate die DNA Konzentrationen gemessen, um zu sehen, in welcher Suspension sich die DNA befand. Die Ergebnisse sind in Tabelle 2 zusammengefasst. Wie aus den Daten ersichtlich, hat die Nachreinigung der DNA mit der Methode der Firma Zymo besser funktioniert als die der Firma Machery und Nagel. Obwohl bei der ersten Methode fast die Hälfte der DNA durch diesen Schritt verloren ging, wurde bei der zweiten Methode die DNA Menge auf ca. 1/10 reduziert. Die Nachreinigung der DNA mit Ampurebeads funktionierte am besten bei *B. burgdorferi* isolate B31, da hier fast die gesamte DNA Menge erhalten blieb (28,4 ng/µl von ursprünglich 29,8 ng/µl) während bei der Verwendung von Lambda DNA das Zymo Kit am besten funktionierte (3,6 ng/µl von ursprünglich 4,9 ng/µl). Die Verwendung von Ampurebeads mit Polyethyleneglycol (PEG) zeigte das schlechteste Ergebnis.

Die Daten zeigen, dass nicht alle als geeignet beschriebenen Methoden es auch sind, und dass Methoden unbedingt ausgetestet werden müssen, bevor sie an wertvollen Proben angewendet werden.

**Tabelle 2:** Reinigung von extrahierter DNA mit unterschiedlichen Methoden

DNA messung im Qubit Photometer

DNA	Ausgangs DNA Konz ng/μl	Zymo ng/μl	MN <sup>a</sup> ng/μl	Ampure beads (Agilent)			Ampure beads (mit PEG solution)*		
				DNA + Beads	ÜB <sup>b</sup> ng/μl	Eluat ng/μl	DNA+Beads	ÜB <sup>b</sup> ng/μl	Eluat ng/μl
B31-147/9	29.8	14.5	3.48	1+1	2.7	28.4	1+0.7 <sup>c</sup>	2.9	0.206
							1+0.7 <sup>d</sup>	1.49	4.14
Lamda 10 <sup>-1</sup>	4.86	3.06	0.67	1+1	0.386	2.5	1+1	3.36	0.18
				1+0,5	0.506	2.08	1+0,5	4.58	0.186

<sup>a</sup>MN=Machery und Nagel

<sup>b</sup>ÜB=Überstand

<sup>c</sup>Verdünnung Beads in Puffer 1:250

<sup>d</sup>Verdünnung Beads in Puffer 1:25

\*Ramawatar und Schwessinger, DNA size selection (>3-4 kb) and purification of DNA using an improved homemade SPRI beads solution.

[dx.doi.org/10.17504/protocols.io.n7hdhj6](https://doi.org/10.17504/protocols.io.n7hdhj6) (Methode siehe Appendix)

### 2.1.1.2 Magnetobeads unterschiedlicher Hersteller

Im Weiteren wurden Methoden, die beide Magnetobeads verwenden und von den Firmen Promega und AnalytikJena angeboten werden, verglichen. Die Methoden der beiden Firmen unterscheiden sich im Hinblick auf eingesetzte Puffer sowie in der Inkubationszeit der Proben. Während bei der Firma Promega eine Inkubationszeit mit Lysepuffer und Proteinase K zwei Stunden dauert, nimmt der Lyse- und Proteinverdauungsschritt bei Analytik Jena nur eine Stunde in Anspruch.

Die DNA Aufreinigung wurde mit Borreliensuspension aus der gleichen Kultur und mit der gleichen Anfangsmenge (Ausgangsmaterial 50µl Zellsuspension) mit beiden Verfahren durchgeführt. Um zu sehen, wie sich geringe Zellmengen (d.h. geringe DNA Mengen im Ansatz) auf den Erfolg der verschiedenen Methoden auswirkte, wurde eine Verdünnungsreihe der ursprünglichen Zellsuspension (jeweils 1:10) unter gleichen Bedingungen aufgereinigt. Die folgenden Kits wurden verwendet: Maxwell 16LEV Blood DNA Kit; Analytika Jena: innuPREP Bacteria DNA Kit-IPC 16. Alle Proben wurden mit 200µl Elutionspuffer des jeweiligen Kits eluiert bzw. wenn andere Mengen zur Elution notwendig waren, wurde das Eluat nachträglich auf 200 µl aufgefüllt. Die Effizienz der DNA Aufreinigung wurde mit Hilfe einer real-time PCR (Strube, Montenegro et al. 2010) getestet. Zielgen dieser PCR ist das bei Lyme borreliose Spezies duplizierte 23S rRNA Gen. Alle Verdünnungen wurden in einem Ansatz aufbereitet und die Ct Werte im Stratagene real-time PCR-Cycler gemessen. Der Ct Wert gibt Auskunft über die Menge an DNA, die in den Proben vorhanden ist. Je größer der Ct Wert, umso geringer die Menge an DNA. Erfahrungsgemäß ist bekannt, dass ein Ct Wert Unterschied von 2-4 einer ca. 10fachen Verdünnung der Probe entspricht.

**Tabelle 3** Vergleich von Maxwell und Analytika Jena DNA Aufreinigung mittels Magentobeads. Borrelienstamm PHei 80/16 in verschiedenen Verdünnungen

Zellenzahl gezählt	Qubit ng/μl# (konz) <sup>§</sup>	2*10 <sup>8</sup>	2*10 <sup>8</sup>	2*10 <sup>7</sup>	2*10 <sup>7</sup>	2*10 <sup>6</sup>	2*10 <sup>6</sup>	2*10 <sup>5</sup>	2*10 <sup>5</sup>
Verdünnung		konz	konz	10 <sup>-1</sup>	10 <sup>-1</sup>	10 <sup>-2</sup>	10 <sup>-2</sup>	10 <sup>-3</sup>	10 <sup>-3</sup>
<b>A) Probenvorbereitung wie vom Gerätehersteller vorgegeben</b>									
Probenaufber.: Maxwell Gerät: Maxwell	0,504	Ct 17,30	Ct 17,94	Ct 21,16	Ct 21,34	Ct 25,10	Ct 25,68	Ct 28,92	Ct 28,87
Probenaufber.: innu prep IPC 16 Gerät: Analytika	0,160	Ct 19,29	Ct 19,03	Ct 23,20	Ct 23,06	Ct 26,90	Ct 28,60	Ct 31,57	Ct 30,33
<b>B) Probenvorbereitung des jeweils anderen Gerätehersteller</b>									
Probenaufber.: innu prep IPC16 Gerät: Maxwell	0,314	Ct 17,64	Ct 17,75	Ct 21,93	Ct 22,81	Ct 26,37	Ct 25,72	Ct 29,53	Ct 29,33
Probenaufber.: Maxwell Gerät: Analytika	zu gering	Ct 20,88	Ct 20,81	Ct 24,59	Ct 23,83	Ct 29,97	Ct 29,31	Ct 31,99	Ct 32,32

#Qubit STD 1 39,08; Qubit STD 2 12730,82. Die Konzentration der DNA wurde nach der Aufreinigung gemessen  
<sup>§</sup> Ausschließlich für die konzentrierte DNA Präparation konnten Qubit Werte gemessen werden, in den anderen Verdünnungen war nicht genügend DNA vorhanden, um photometrische Messungen zuverlässig durchzuführen.

Wie aus Tabelle 3 ersichtlich, ergab die Aufreinigung mit dem Maxwell Gerät bessere Ergebnisse als die Aufreinigung mit dem Gerät der Firma Analytik Jena.

Interessanterweise wurden, auch wenn die Probenvorbereitung von Analytik Jena verwendet wurde, mit dem Maxwell fast gleich gute Ergebnisse erzielt wie mit Verwendung des Gesamtprotokolls des Maxwell. Das spricht dafür, dass die von Promega eingesetzten Kartuschen zur DNA Aufreinigung robuster ist und weniger empfindlich auf die eingesetzte Chemie reagiert. Bei beiden Versuchsanordnungen lagen die Ct Werte im gleichen Bereich (Tabelle 3A und 3B). Im Vergleich dazu lagen die Ct Werte bei Verwendung des Analytik Jena Gerätes jeweils 2-3 Ct Werte höher als bei Verwendung des Promega Maxwell Gerätes. Auffällig war, dass bei Verwendung der

Promega Chemikalien und des zeitlich längeren Lyseverfahrens, bei der Aufreinigung mit dem Gerät der Firma Analytik Jena keine besseren Ergebnisse erzielt wurden; auch hier lagen die Ct Werte höher als bei Verwendung des Maxwell Gerätes. Das könnte auf Inkompatibilität von Lysepuffer und Gerätechemie zurückzuführen sein. Wir haben mit der Firma Analytik Jena Kontakt aufgenommen, um herauszufinden, worauf diese Unterschiede möglicherweise zurückzuführen sind.

### 2.1.1.3 Magnetobeads versus SMART Spitzen

Die Firma AnalytikJena hat ein Verfahren entwickelt, das zur Extraktion von qualitativ besonders hochwertiger und besonders langkettiger DNA führen soll. Darüber hinaus wirbt die Firma damit, dass dieses Verfahren zu quantitativ hervorragenden Ergebnissen führen soll. In diesem Protokoll werden mit einer Matrix beschichtete Spitzen verwendet, welches sich auch besonders gut für gering konzentrierte DNA eignen soll (Zitat der Firma: „Because of the extremely high binding capacities of the modified surfaces used, the achieved yield of nucleic acids is practically unlimited, as opposed to magnetic particle based extraction, for example. In addition, the new technology also allows the isolation of nucleic acids that are only present in very low concentrations in a biological sample.“). Diese Methodik wurde von der Firma Analytik Jena als “SMART Extraktion” bezeichnet.

In den folgenden Versuchen wurden wiederum Verdünnungsreihen von Borrelien verwendet. Als Ausgangsmaterial für alle Versuche wurden 50µl Zellsuspension verwendet. Alle Proben wurden mit 200µl eluiert bzw. falls andere Elutionsvolumina erforderlich waren, wurde nach Beendigung des Aufreinigungsverfahrens das Elutionsvolumen auf 200µl aufgefüllt. Extrahierte DNA von allen Verdünnungsstufen wurde in einem Ansatz aufbereitet, für die real-time PCR vorbereitet und im Stratagene Cyclus gemessen. Es wurde wiederum das Protokoll der Firma Promega (Aufreinigung im Maxwell Gerät) mit den Protokollen der Firma Analytik Jena (Aufreinigung mittels Magnetobeads mit InnuPREP und mittels SMART Spitzen) verglichen.

Wie die Ergebnisse in Tabelle 4 zeigen, schnitt auch in diesem Versuch das Protokoll der Firma Promega am besten ab. Wie zuvor, war die Magnetpartikel Aufreinigung mit dem Protokoll der Firma Analytik Jena etwas schlechter, alle Ct Werte

lagen ca. 2 Stufen höher als bei Maxwell. Die Ct Werte der SMART Aufreinigung lagen am höchsten, ca. 2-4 Stufen über den Werten der Magnetpartikelreinigung. Für die letzte Verdünnungsstufe konnten keine Ct Werte ermittelt werden.

Eine photometrische Messung der DNA nach Aufreinigung war mit dem Qubit in den Proben nicht möglich, da ungenügende Mengen von DNA vorhanden waren, um solche Messungen zuverlässig durchzuführen.

**Tabelle 4** Validierung zwischen Maxwell (Magnetpartikel), Analytik Jena (Magnetpartikel, innuPREP) und Analytik Jena (SMART prep) mit WudIIStamm in verschiedenen Verdünnungen

Zellenzahl gezählt	10 <sup>7</sup>	10 <sup>7</sup>	7*10 <sup>5</sup>	7*10 <sup>5</sup>	5*10 <sup>4</sup>	5*10 <sup>4</sup>	10 <sup>3</sup>	10 <sup>3</sup>
Geräte	WudII konz	WudII konz	WudII 10 <sup>-2</sup>	WudII 10 <sup>-2</sup>	WudII 10 <sup>-3</sup>	WudII 10 <sup>-3</sup>	WudII 10 <sup>-5</sup>	WudII 10 <sup>-5</sup>
<b>Maxwell Qubit</b>	-	-	-	-	-	-	-	-
<b>Maxwell Stratagene Ct Werte</b>	18,03	17,27	26,22	26,33	30,38	30,98	33,58	34,79
	Resuspension Trispuffer							
<b>innuPREP B.DNA IPC16 Qubit</b>	-	-	-	-	-	-	-	-
<b>innuPREP B.DNA IPC16 Stratagene Ct</b>	19,51	19,90	28,00	28,22	31,78	-	34,90	35,75
	Resuspension Trispuffer							
<b>Smart DNA prep Qubit</b>	-	-	-	-	-	-	-	-
<b>Smart DNA prep Stratagene Ct</b>	20,10	20,84	33,48	32,77	35,78	35,81	No Ct	No Ct

### 3 Vergleich unterschiedlicher Sequenziermethoden

Im folgenden Abschnitt wird nochmals auf die unterschiedlichen Sequenziermethoden eingegangen und diese im Vergleich dargestellt. Besonders wird auf die Nanopore Technologie eingegangen, da sie im Laufe dieses Projektes am LGL etabliert und getestet wurde.

#### 3.1 Illumina Sequenzierung

Die Methode der Illumina MiSeq Sequenzierung ist vor einigen Jahren primär vom NRZ für Borrelien am LGL etabliert worden und wird zur Sequenzierung bakterieller Genome sowie neuerdings auch zur Sequenzierung von MLST PCR Produkten angewendet. Vom NRZ für Borrelien wurden seit der Beschaffung des Gerätes in 2013 mehr als 100 Genome sequenziert und auch neue Spezies beschrieben, z.B. *Borrelia maritima* (Manuskript eingereicht bei IJSEM).

Es hat sich gezeigt, dass sich mit Illumina Sequenzen große Teile des Kerngenomes zusammensetzen lassen, die für vergleichende Untersuchungen bakterieller Isolate wunderbar herangezogen werden können (z.B. für die Erstellung von Stammbäumen oder CoreMLST), dass aber Genomregionen mit langen repetitiven Sequenzen (wie z.B. das *Imp1* Gen im Borrelienchromosome) oder auch Borrelien Plasmide sich nicht oder nur unvollständig zusammensetzen lassen (Becker, Margos et al. 2016; Castillo-Ramirez, Fingerle et al. 2016; Margos, Hepner et al. 2017; Marosevic, Margos et al. 2017; Tyler, Tyson et al. 2018). Nichtsdestotrotz sind Illumina Sequenzen aufgrund ihrer Akuratheit hervorragend geeignet (oder sogar unabdingbar) im Zusammenspiel mit langkettigen Sequenziermethoden bei der Zusammensetzung von Gesamtgenomen zu helfen und Fehler in der Basenfolge auszumerzen. Bedingt durch die relative Einfachheit der Libraryproduktion, die sehr geringen Mengen an benötigter DNA sowie die sehr große Genauigkeit der Sequenzen, bietet sich der Illumina MiSeq für Betreiber an, bei denen nicht übermäßig viele Sequenziermengen anfallen, die Sequenzierung großer Genome oder die Sequenzierung von Gesamtgenomen nicht im Vordergrund stehen bzw. durch Voruntersuchungen schon gut definiert sind (Abb 1). Im Rahmen des hier beschriebenen Projektes wurden bereits vorhandene Illumina Sequenzen von Borrelienisolaten genutzt, um unterschiedliche Software zur Genomassemblierung zu eruiieren.

#### 3.2 Pacific Bioscience

Die Sequenzierung mit Pacific Bioscience Technologie wurde von der Genomics Core Facility der Universität Oslo durchgeführt. Diese Technologie erfordert das Betreiben großer und sehr teurer Geräte, was sich eher für professionelle, auf Genomics ausgerichtete Einrichtungen lohnt. Wie bereits eingangs erwähnt, erforderte die SMRT

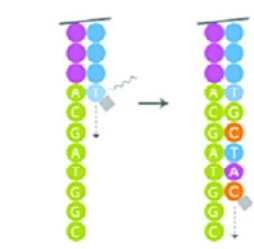
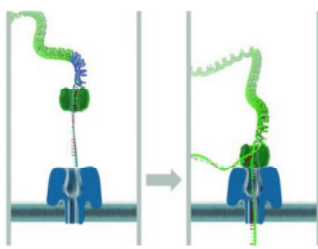
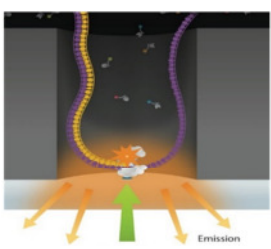
Technologie sehr große Mengen an DNA (10 microgramm), was damit zusammenhängt, dass eine Größenselektionsschritt durchgeführt wird, bei dem großen Mengen an DNA verlorengehen können. Darüber hinaus war es bis Anfang dieses Jahres nur möglich, einzelne Bakterienisolate pro FlowCell laufen zu lassen, da die Möglichkeit des Multiplexens nicht gegeben war. Daher waren SMRT Sequenzen im Vergleich zu Illumina sehr teuer, die Kosten beliefen sich auf ca. 1.200 – 1.500 Euro pro Isolate. Dies hat sich seit Anfang dieses Jahres geändert, da eine neue Sequenzierchemie entwickelt wurde und die Möglichkeit des Multiplexens nun zur Verfügung steht. D.h., es können nun ca. 16 Borrelienisolate auf einer FlowCell sequenziert werden, was die Kosten pro Isolate auf ca. 400 Euro gesenkt hat. Weiterhin ist dadurch der Bedarf an genomischer DNA gesunken und es werden ‚nur‘ noch ca. 2-3 microgramm an genomischer DNA pro Isolate benötigt.

Die bisher gemachten Erfahrungen mit SMRT Sequenzierung haben gezeigt, dass diese Technologie sich außerordentlich gut eignet, um Borreliengenome vollständig zusammensetzen (Margos, Hepner et al. 2017). Der Stamm B31 wurde vom NRZ für Borrelien mittels unterschiedlicher Technologien sequenziert (Illumina, SMRT, Nanopore) und mit Referenzgenomen, die in der NCBI GenBank hinterlegt sind verglichen. Dabei hat sich gezeigt, dass SMRT Technologie in Verbindung mit Illumina die besten Ergebnisse brachte, da das Hauptchromosom und viele der Plasmide nur ganz geringe Unterschiede zur GenBank Referenz aufwiesen.

### 3.3 Nanopore Sequenzierung und Etablierung am LGL

Nanopore Sequenzierung ist vor relativ kurzer Zeit auf den Markt gekommen und war zu der Zeit – verglichen mit den anderen Methoden – noch nicht voll ausgereift. Obwohl mit dieser Methode ultralange Reads erstellt werden können, ist die Fehlerrate doch beträchtlich. Die Menge an DNA, die benötigt wird, um sequenzierfähige Libraries herzustellen, liegt im Bereich zwischen Illumina und Pacific Bioscience SMRT Technology. Die apparative Voraussetzung für Nanopore Technologie ist sehr viel einfacher als für die beiden anderen Technologien und die Anwendung kann sehr flexibel gestaltet werden.



ILLUMINA	OXFORD NANOPORE	PACIFIC BIOSCIENCES
		
Sequenzierung mittels Synthetisierung	Strang-Sequenzierung	Strang-Sequenzierung mittels Synthetisierung
Messung von basenspezif. Fluoreszenzsignalen	Messung der basenspezif. Stromänderung	Messung von basenspezif. Fluoreszenzsignalen
0,2-100 ng DNA	1 µg DNA	10 µg DNA
kurze Reads (50-600 bp)	ultralange Reads (1-100 kb)	lange Reads (10-20 kbp)
äußerst akkurat	höhere Fehlerrate	höhere Fehlerrate

**Abbildung 1:** Vergleich der unterschiedlichen NGS Methoden

(Rhoads and Au 2015; de Lannoy, de Ridder et al. 2017)

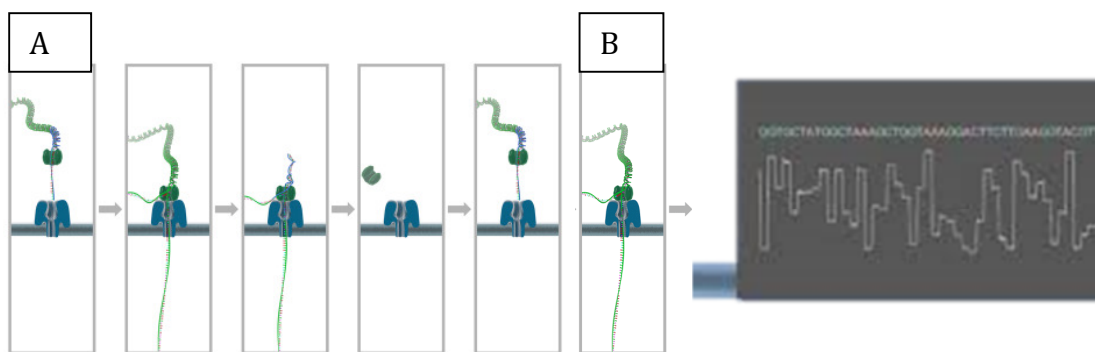
([https://www.illumina.com/documents/products/illumina\\_sequencing\\_introduction.pdf](https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf))

Um die Nanopore Technologie am LGL zu etablieren, wurden zunächst die apparativen Voraussetzungen geschaffen. Hierzu wurde ein Computer mit genügend hohen Speicherkapazitäten beschafft, und die entsprechenden Softwares für die Kommunikation mit dem MinION, für Basecalling und zur Bearbeitung der Sequenzen installiert. Das kleinste Gerät der Firma ONT, der MinION, wurde mit dem Starterpack bei der Firma Nanopore beschafft.

Zum Betreiben des MinION (Abb. 2) muss die von der Firma Nanopore zur Verfügung gestellte Software MinKNOW auf dem Computer für die Verbindung zum Gerät und zur Steuerung der Sequenzierung installiert sein. Darüber hinaus ist noch die Software ‚Albacore‘ erforderlich, die für die Umwandlung der gemessenen Spannungsdifferenzen in eine Basenabfolge (base calling) notwendig ist (Abb. 3). Diese Voraussetzungen wurden problemlos mit Hilfe der Firma und unserer EDV geschaffen.



**Abbildung 2:** MinION tragbarer Sequenzierer im Taschenformat, der mit einem Computer verbunden werden kann. Die 512 Nanoporekanäle können mit jeweils vier Nanoporen kommunizieren.



**Abbildung 3:** Nanopore Sequenzierung. Die Graphiken A und B zeigen, wie der DNA Strang durch die Nanopore hindurchgezogen wird (A) und die Signalerfassung (B).

([https://github.com/aechchiki/SIB\\_LongReadsWorkshop\\_Zurich17/wiki/Oxford-Nanopore-Technologies:-MinION](https://github.com/aechchiki/SIB_LongReadsWorkshop_Zurich17/wiki/Oxford-Nanopore-Technologies:-MinION); [http://www.nucleomics.be/platforms/oxford-nanopore-gridion-system/.](http://www.nucleomics.be/platforms/oxford-nanopore-gridion-system/))



**Abbildung 4:** FlowCell in einem MinION (A) und Schema der Nanoporenverteilung (de Lannoy, de Ridder et al. 2017). (A) zeigt die FlowCell im MinION, Der Bereich, in dem die DNA Sequenzierung abläuft, ist mit einem Pfeil markiert. (B) zeigt die Anordnung der Nanoporen; die größeren bunten Kreise sind die Wells in denen sich die Nanoporen befinden (kleine Kreise). Nur Wells die intakt sind und nur eine Nanopore enthalten, sind funktionsfähig.

In dem Gerät sind 512 Nanoporekanäle vorhanden, von denen jeder mit vier Nanoporen verbunden ist (Mux1-4) und deren Daten beim Sequenzierlauf erfassen kann (Abb. 4A). Das bedeutet, dass auf jeder FlowCell über 2,000 Nanoporen vorhanden sind, über die die Sequenzierung ablaufen kann. Nicht alle Nanoporen sind funktionsfähig (Abb. 4B), deshalb stehen oft ‚nur‘ 1.200-1.500 Poren zur Verfügung. Die Firma empfiehlt, FlowCells, die weniger als 800 sequenzierfähige Poren hat, für eine Ersatzlieferung zurückzuschicken. Dies kann über einen FlowCell-check getestet werden, welcher innerhalb von 10 Tagen nach Erhalt durchgeführt werden muss. FlowCells können bis zu 48 Stunden laufen und dabei können 10-30 Gb Sequenzdaten generiert werden. Es steht eine Vielzahl von Kits für die Produktion von Libraries zur Verfügung, die mit oder ohne Fragmentierung der DNA, mit oder ohne PCR, mit Barcode benutzt werden können, und die unterschiedliche Mengen an genomischer DNA benötigen. Die im Rahmen dieses Projektes verwendeten Kits sind in Tabelle S1 im Anhang1 zu finden.

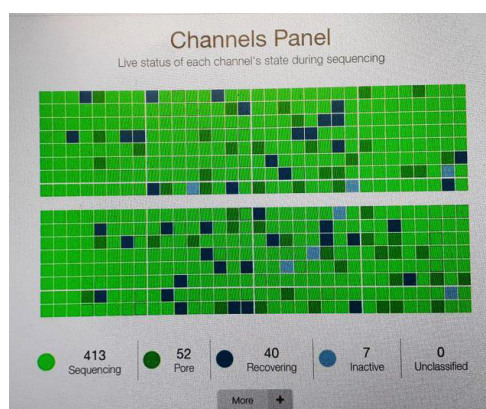
### 3.3.1 Sequenzierläufe

Wie von der Firma Nanopore empfohlen wurde zunächst ein FlowCell-check durchgeführt. Beide FlowCells hatten >1,000 Nanoporen, die zur Sequenzierung geeignet waren, und lagen somit über dem von ONT angegebenen Wert von >800 Nanoporen. Danach wurde – ebenfalls wie von ONT empfohlen – erstmal mit Lambda

DNA gearbeitet. Zunächst wurde das Kit SQK-LSK108 verwendet. Für die Libraryproduktion wurden exakt die von ONT angegebenen Schritte eingehalten. Generell werden hier hydrophobe Adapter an die DNA Stränge ligiert sowie ein Motorprotein, das die DNA durch die Nanopore hindurchschieben kann (Abb. 3A). Die in der Nanopore befindlichen DNA Basen sorgen für eine Veränderung des Spannungspotentials, welches erfasst wird (Abb. 3B) und später in Basensequenzen umgewandelt werden kann.

Für die erste Präparation von Lambda DNA wurden normale 1.5 ml Eppendorf tubes verwendet. Eine DNA Fragmentierung für dieses Kit ist optional und wurde für die erste DNA Präparation nicht angewendet, damit längere DNA Fragmente zur Sequenzierung zur Verfügung standen. In diesem Falle wurde die DNA Menge während der Libraryproduktion nicht erfasst. Während des Laufs zeigt die Software verschiedene Parameter an, mit denen die Qualität des Laufs beurteilt werden kann (Abb. 5).

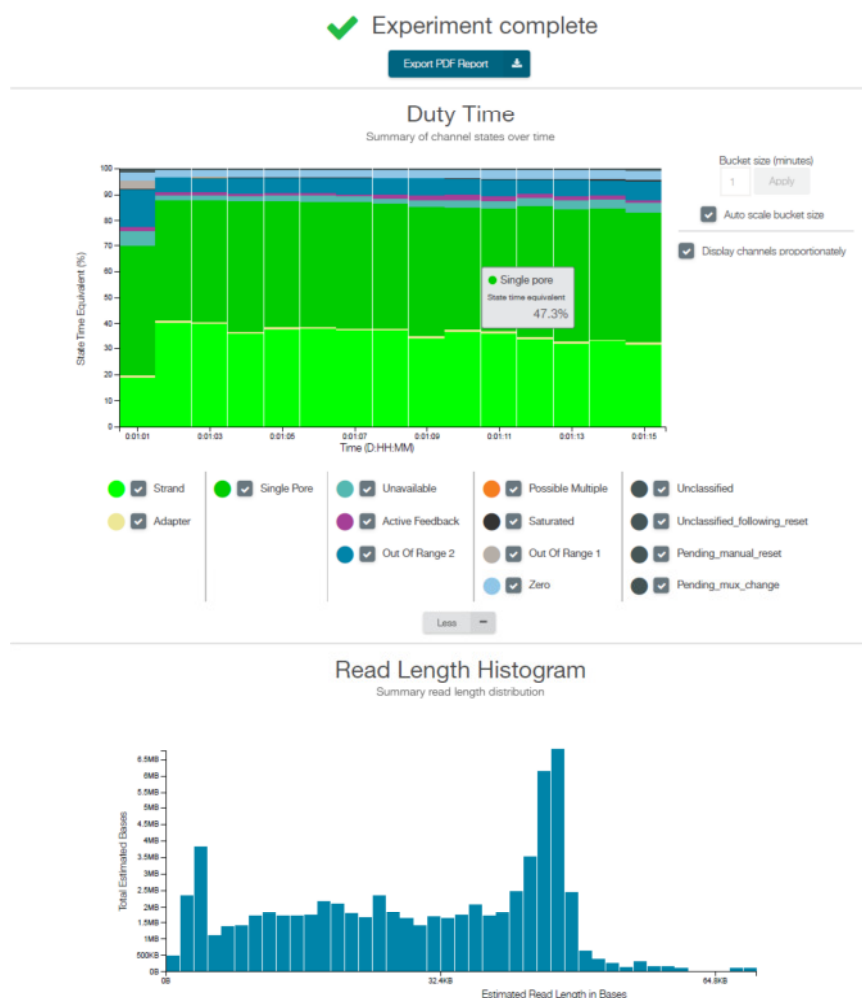
Nachdem der Sequenzierlauf gestartet war, stellte sich heraus, dass die DNA während der Libraryproduktion stark abgenommen hatte. Dies war offensichtlich, da bei diesem Lauf das Verhältnis von sequenzierenden Pores : sequenzierfähigen Poren sehr schlecht war, es lag bei 1:1 (sollte präferentiell 3:1 oder 2:1 sein), deshalb wurde der Lauf nach kurzer Zeit abgebrochen. Dieser Versuch zeigte eindrucksvoll, dass normale Eppendorf Gefäße nicht zur Herstellung von DNA Libraries für Nanopore geeignet sind, da zu viel DNA an die Gefäße gebunden wurde, und unbedingt DNA low-bind Eppendorf Gefäße verwendet werden sollten.



**Abbildung 5:** Während des Nanoporelaufs wird die Anzahl von sequenzierenden Poren (,Sequencing'), sequenzierfähigen Poren (,Pore'), recovering Poren (,Recovering') und inaktiven Poren (,inactive') angezeigt. Dies gibt Auskunft über die Qualität des Laufs bzw. die Quantität der geladenen DNA library.

(<https://www.twitur.com/account/Barefootmum/tweet/1096105147338493953>)

Trotzdem war dieser Lauf ausreichend, um festzustellen, dass die Softwareverbindung mit dem MinION gut war und die Laufparameter wurden gut erfasst. Das Genom von Lambda Phagen ist ca. 48 kb und es zeigte sich, dass in der Region zwischen 40 und 48 kb ein Peak war, was auf nicht oder nur gering fragmentierte DNA hinweist. Es gab jedoch DNA Fragmente im gesamten Bereich zwischen 0 und 40 kb, was zeigt, dass es immer zu Fragmentierung der DNA kommt, wahrscheinlich durch Scherkräfte beim Pipettieren. Abbildung 6 zeigt exemplarisch den Laufreport. Reports für weitere Läufe sind im Appendix dargestellt.



**Abbildung 6:** Report Experiment 1, Lambda DNA. Im obersten Feld ist ersichtlich, dass das Verhältnis von sequenzierenden Poren : sequenzierfähigen Poren eher gering war (zu wenig DNA auf FlowCell). In unteren Bereich ist die Länge der sequenzierten DNA Fragmente gezeigt. Der Peak in der Region zwischen 40-48 kb zeigt nicht fragmentierte Lambda DNA (Genomgröße ca. 48 kb).

In einem zweiten Versuch mit Lambda DNA wurden die von der Firma Nanopore empfohlenen DNA low-binding Eppendorf Tubes verwendet und die DNA in g-tubes (Covaris) fragmentiert. Bei Verwendung der g-tubes wurden genau die vom Hersteller angegebenen Zentrifugationsschritte und -geschwindigkeiten eingehalten, um 8 kb Fragmente zu erhalten. Anschließend wurde das gleiche Kit (SQK-LKS108) wie beim ersten Mal verwendet. Der Verlust von DNA wurde im Qubit Photometer nach jedem Schritt gemessen. Wie Tabelle 5 zeigt, ist selbst bei Verwendung von DNA low-bind Tubes mit großen Verlusten von DNA während der Libraryproduktion zu rechnen. Es scheint besonders während des Adapterligationsschrittes zu erfolgen; nach diesem Schritt war ca. nur noch  $\frac{1}{4}$  der Ausgangsmenge vorhanden während es bei den anderen Schritten zu viel geringeren Verlusten gekommen war. Die verbliebene Menge an DNA reichte aber für einen Lauf aus.

**Tabelle 5:** Verlust von DNA während der Libraryproduktion mit DNA low-bind Tubes

<b>Library prep Lambda DNA</b>	<b>DNA Konzentration</b>	<b>Volumen</b>	<b>Gesamt-DNA</b>
Start	64 ng/ $\mu$ l	20 $\mu$ l	1.280 ng
g-tube Fragmentierung	24 ng/ $\mu$ l	45 $\mu$ l	1.000 ng
Endrepair und Ampurebeads	32ng/ $\mu$ l	30 $\mu$ l	960 ng
Adapterligation	27 ng/ $\mu$ l	14 $\mu$ l	378 ng

Bei diesem Lauf lag das Verhältnis sequenzierende Poren : sequenzierfähige Poren 2:1, was wesentlich besser war als zuvor. Dem Report war zu entnehmen (Anhang1\_Abb. S2), dass es zwischen 6 und 10 kb einen Sequenzpeak gab und nur sehr wenige Reads im niedrigen Bereich von ca. 1 kb zu finden waren.

Der Tabelle 6 sind die in diesem Projekt durchgeführten Librarypräparationen und durchgeführten Sequenzläufe zu entnehmen, die im Folgenden weiter diskutiert sind.

**Tabelle 6:** Isolate für Nanopore Sequenzierung und verwendete Kits

Isolate	Library kit	DNA [ng] Lib.prep	DNA [ng] FloCell load	Runtime	Remarks	Result/Bemerkungen
Lambda DNA-1	Ligation Sequencing Kit (SQK-LKS108)	nd	nd	10 Minuten	S:P 1:1	Verlust von DNA; zu wenig DNA auf FlowCell;
Lambda DNA-2	SQK-LKS108	1.280	378	10 Minuten	S:P 2:1	gut
Fem19	SQK-LKS108	1.348	292	3 h	S:P 0.3:1	nicht genügend DNA; sehr schlechter Lauf
B31 138-8	SQK-LKS109*	1.125	354	20 h	S:P 2:1	gut
<i>Borrelia turcica</i> MLST PCR Produkte	Native Barcoding Kit (EXP-NBD104 mit SQK-LKS108)	nd	nd	4 h	S:P 0.5:1	hier wurden PCR Produkte verwendet wurden, deshalb war die Menge an DNA nicht entscheidend
BR149, BR151, BR208, BR193, Coryne	PCR Barcoding Kit (SQK-PBK004)	100 ng	43, 107, 308, 96, 152	16 h	S:P 0.3:1	Genügend Sequenzierdaten ;

\*SQK-LKS109 ist ebenfalls ein Ligation Sequencing Kit; der Unterschied zu LKS108 besteht darin, dass mit LKS109 am Ende der Libraryproduktion ein Größenselektionsschritt eingefügt werden kann.

Obwohl die Libraryproduktion für Isolate Fem19 ganz gut war, wurden trotzdem keine oder nur ungenügende Sequenzen erhalten. Grund dafür war, dass eine Luftblase in die Sequenzierkammer eingebracht wurde, die eine gute Sequenzierung verhinderte. Der Eintrag von Luftblasen auf die Sequenzierkammer zerstört die Integrität der FlowCell und kann – je nach Größe der Luftblase – dazu führen, dass die FlowCell unbrauchbar wird. Den Daten war jedoch zu entnehmen, dass ein Sequenzpeak im Bereich von 1-2 kb erhalten wurde. Frühere Sequenzierungen (durchgeführt am Gene Center der LMU) hatten gezeigt, dass es zu einer starken Sequenzierung von solchen nicht erwünschten ‚kleinen‘ DNA Fragmenten kommen kann und für weitere Läufe wurde daher das Kit SQK-LSK109 verwendet, da hier eine Größenselektion am Ende der Libraryproduktion stattfinden kann.

Deshalb wurden im Weiteren zunächst Versuche mit dem *Borrelia burgdorferi* Stamm B31 durchgeführt, da es für diesen Stamm gute Referenzsequenzen in der NCBI Datenbank (GenBank) (Fraser, Casjens et al. 1997) und durch eigene Arbeit gibt (Margos et al. 2017). Die DNA wurde mittels g-tubes auf 20 kb fragmentiert und die Library wie angegeben produziert. Obwohl ein Größenselektionsschritt eingefügt wurde, mit dem angeblich DNA Fragmente <3000 bp eliminiert werden sollen, war dies nicht der Fall. Aus dem Sequenzierreport (Anhang1\_Abb. S3) ist zu erkennen, dass es zwei Sequenzpeaks gibt: einen etwas größeren bei ca. 10 kb, den anderen im Bereich von 1-2 kb. D.h. durch die Größenselektion konnten nicht alle kleinen DNA Fragmente <3000 bp herausgefiltert. Die Anwesenheit von diesen kleinen DNA Fragmenten (1000-2000 bp) wirkt sich negativ auf die Genom Assemblierung aus, deshalb wurden Versuche unternommen (siehe Verbesserung der Libraries). Dieses Thema wird und wurde in den Nanopore ‚Community‘ diskutiert und es gibt eine Reihe von Protokollen, die im Rahmen dieses Projektes ausprobiert wurden (siehe Pkt. 2.1.1. DNA Aufreinigungsverfahren). Wie bereits oben erwähnt, waren diese Methoden nicht wirklich hilfreich bei der Eliminierung von kurzen DNA Fragmenten, deshalb sollte es schon bei der Extraktion vermieden werden, solche Fragmente zu produzieren, evtl. durch Verwendung von weiten Pipettenspitzen, die von verschiedenen Spitzenherstellern angeboten werden.

### 3.3.2 Sequenzenauswertung

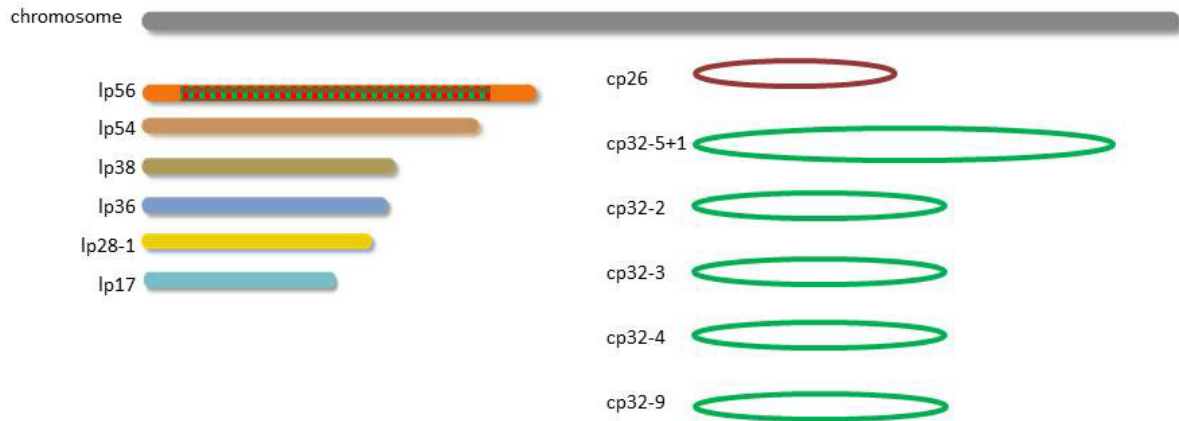
Die Auswertung der Sequenzen und die Zusammensetzung der Genom oder einzelner Genomelemente (z.B. Chromosom, Plasmide) ist entscheidend für die Verwendung von NGS. Es hängt natürlich immer von der Fragestellung ab, die beantwortet werden soll, ob die Zusammensetzung kompletter Genome gefordert ist. Andererseits ist natürlich zu bedenken, dass, je vollständiger ein Genom zusammengesetzt ist, umso besser kann ein Vergleich mit anderen Genomen ausgeführt werden. So reicht es für Ausbruchsuntersuchungen oft aus, nur das Kerngenom zu nutzen, da hier die konservierten Genloci zu finden sind, die für ein Core-MLST herangezogen werden (Jolley 2009; Jolley and Maiden 2010). Wenn es jedoch um die Beurteilung von Pathogenitätsfaktoren oder um Resistenzbestimmung geht, reicht dies oft nicht aus, da solche Elemente oft im Nichtkern-Bereich des Genoms zu finden sind. Die hier ausprobierten Methoden sind sicherlich geeignet,



Im Rahmen dieses Projektes wurden die folgenden Softwares angewendet, um zu sehen, inwieweit das *B. burgdorferi* B31 Genom vollständig zusammengesetzt werden kann:

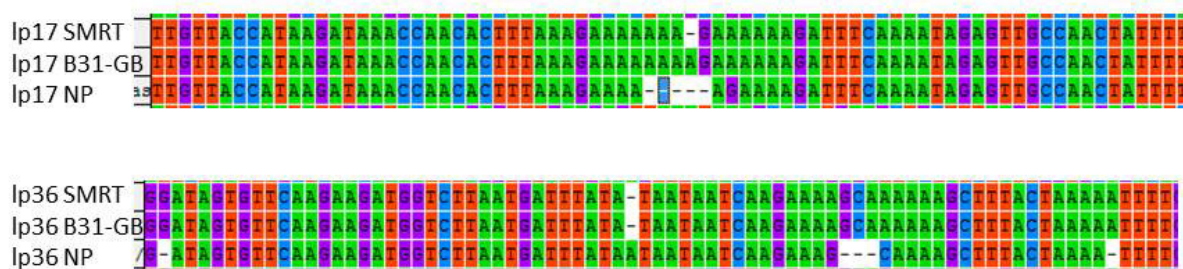
- a) CLC Genomic Workbench, lizenzierte Software von Qiagen
- b) SPAdes version 3.11.1, freie Software, individuelle short- und long-read Assemblies sowie Hybridassemblies (Bankevich, Nurk et al. 2012)
- c) Flye, freie Software, long-read assembly (Kolmogorov, Yuan et al. 2019)
- d) Canu version 1.8, freie Software, long-read Assembly (Koren, Walenz et al. 2016)

Dabei wurde zum Teil auf vorhandene Daten zurückgegriffen. Es muss hier noch betont werden, dass die Firma Pacific Biosciences ihre eigene Software, genannt HGAP (Hierarchical Genome Assembly Process; siehe <https://github.com/PacificBiosciences/Bioinformatics-Training/wiki/HGAP>), zum Zusammensetzen der Genome benutzt. Für die hier verglichenen Pacific Bioscience Daten wurde das Assembly von der Universität Oslo als Teil des Auftrags durchgeführt. Das *B. burgdorferi* Genom wurde ausgewählt um Proof-of-Principle Daten zu erhalten, weil es ein hochkomplexes Genome ist (Abb. 7), in dem viele Plasmide vorhanden sind, die sich z.T. in der DNA Sequenz sehr ähnlich sind. Eine Software, die in der Lage ist, das Borreliengenom akkurat zusammensetzen, kann dies mit hoher Wahrscheinlichkeit auch bei anderen Genomen schaffen.



**Abbildung 7:** Graphik der B31-NRZ Genomeorganisation. Das Genom von B31-NRZ wurde generiert aus Pacific Bioscience SMRT und Illumina MiSeq Konsensussequenzen (Margos et al. 2017a, b). Dafür wurde das SMRT Assembly als Reference in die CLC Genomics Workbench geladen und ein Readmapping mit Illumina MiSeq Sequenzen durchgeführt. Gleiche Farbe bedeutet Zugehörigkeit zur gleichen Plasmidfamilie, lp56 hat eine Insertion von einem cp32 Plasmid. Das resultierende Genom unterschied sich in Hauptchromosom und den meisten Plasmids in wenigen SNPs von der NCBI GenBank Referenz von B31 (Casjens et al. 2012). In den meisten Plasmiden wurden 0-7 SNPs gefunden, das Hauptchromosom wies 45 SNPs auf und lp28-1 179 SNPs. Ausnahme war cp32-5+1, welches über 3.000 Unterschiede zu den entsprechenden Plasmiden in B31-GB aufwies.

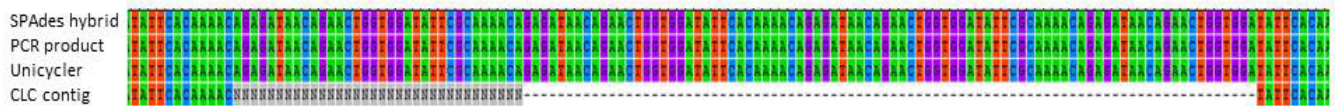
Im Weiteren sind die Ergebnisse von Genomassemblierungen von B31 dargestellt. Abbildung 8 zeigt einen partiellen Vergleich von Zusammensetzungen zweier Plasmide (lp17 und lp36) von SMRT und ONT Sequenzen verglichen mit der B31-GB Referenz. Es ist ersichtlich, dass beide Langreadmethoden Baseninsertionen oder Deletionen im Vergleich zur Referenz zeigen. Diese Fehler können durch Illuminasequenzen ausgemerzt werden.



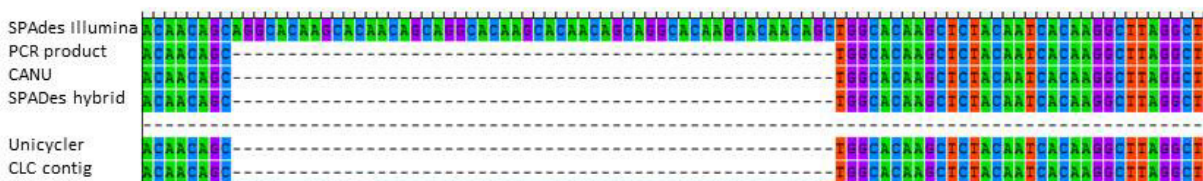
**Abbildung 8:** Vergleich von SMRT und Nanopore Sequenzen mit der *B. burgdorferi* B31 GenBank Referenz (Fraser, Casjens et al. 1997; Casjens, Mongodin et al. 2012).

In der nächsten Abbildungen sind Beispiele aufgeführt, die zeigen, dass jeder Assembler Vor- und Nachteile hat, dass eine Kombination aus langen und kurzen Reads unbedingt notwendig ist, um akkurate Genome zusammzusetzen und dass bisher SMRT in

Kombination mit Illumina MiSeq die besten Ergebnisse im Falle der Borrelien geliefert hat.

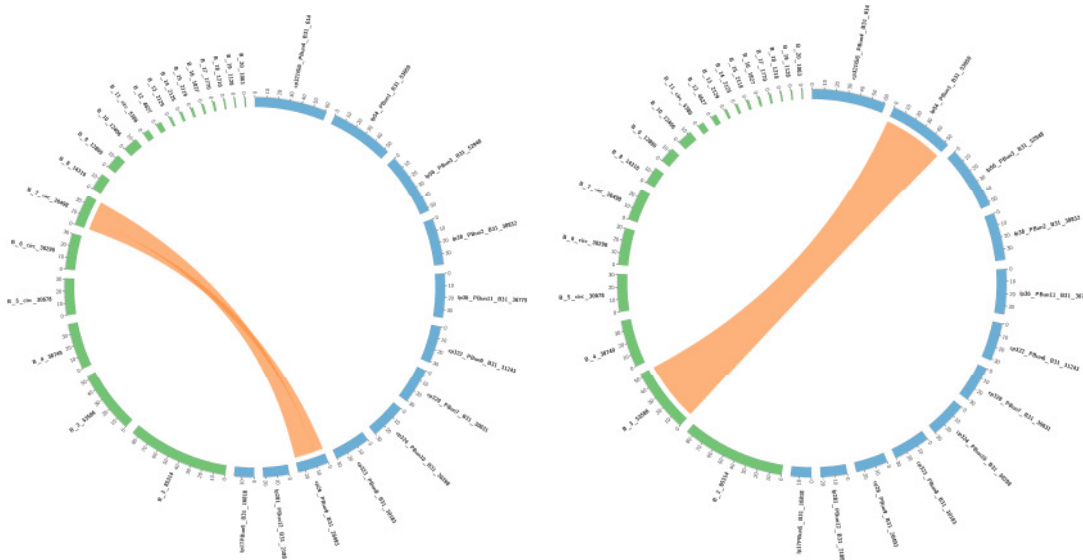


**Abbildung 9:** SPAdes und Unicycler ergaben gleich gute Ergebnisse beim Hybridassembly der chromosomalen repetitiven Region während das Assembly der CLC Genomic Workbench eine erhebliche Lücke aufweist.

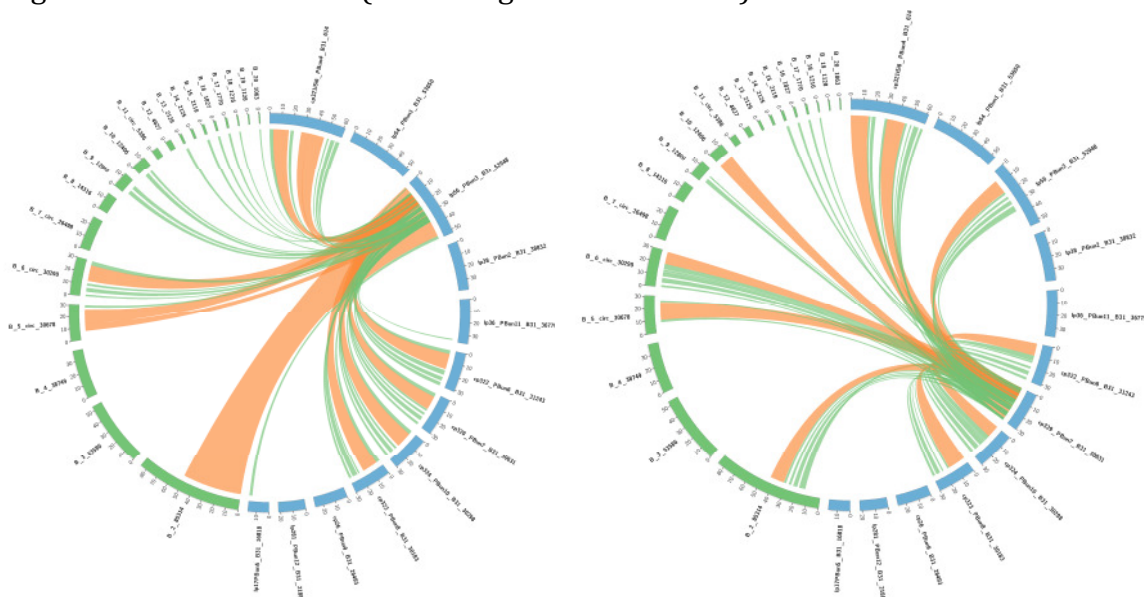


**Abbildung 10:** zeigt eine Region mit repetitiven Sequenzen auf dem Plasmid lp54. Hybridassemblies (SPAdes hybrid, Unicycler), Assembly mit langen Reads (Canu) sowie die CLC Genomics Workbench zeigen weniger Repeats als das SPAdes Assembly mit ausschließlich kurzen Illumina Reads (SPAdes Illumina).

In Anhang2 (erstellt von Alexander Vogel) sind Ergebnisse zusammengefasst, die unterschiedliche Sequenziermethoden (SMRT, ONT, Illumina) und unterschiedliche Assemblyprogramme vergleichen. Es ist ersichtlich, dass das Kerngenom (Chromosom, Plasmide lp54 und cp26) von Borrelien im Allgemeinen gut zusammengesetzt wurde, unabhängig von Sequenziermethode oder Assembler. Problematisch wird es bei den restlichen Plasmiden und hier wurden die besten Ergebnisse mit SMRT Technologie erzielt (Anhang2). Die folgenden Abbildungen zeigen auf der linken Seite Nanopore Reads zusammengesetzt mit Canu verglichen auf der linken Seite mit SMRT assemblies.



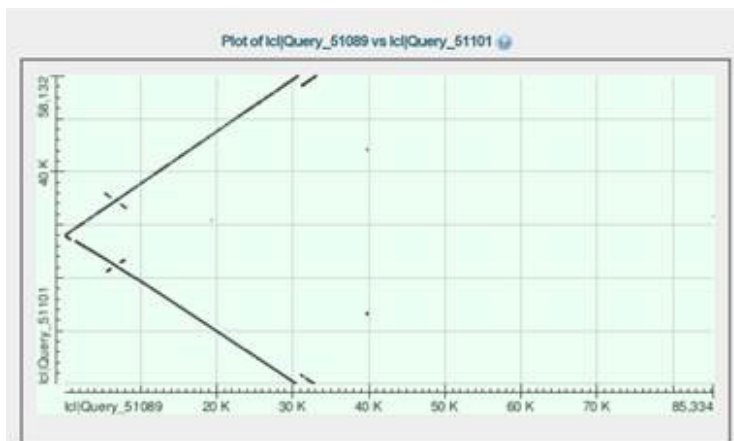
**Abbildung 11:** Vergleich von B31 SMRT (blau) und Nanopore (grün) Sequenzassembly von Plasmiden cp26 (links) und lp54(rechts). Aus dieser Abbildung ist ersichtlich, dass die Plasmide mit beiden Technologien gut sequenziert und beim Assembly gut dargestellt werden können (Abbildung von K. Kuleshov)



**Abbildung 12:** Vergleich von B31 SMRT (blau) und Nanopore (grün) Sequenzassembly von Plasmiden lp56 (links) und cp32-9(rechts). In Plasmid lp56 ist ein cp32 Plasmid integriert und es wird deutlich, dass in allen cp32 Plasmiden homologe Sequenzen zu finden sind (orange und grüne Linien). Die orangen Linien zeigen Sequenzhomologien von >5000 bp, die grünen Linien weisen auf Sequenzhomologien von <5000 bp. Aus dieser Abbildung ist ersichtlich, dass es beim Assembly der cp32 bedingt durch Sequenzhomologien Probleme gibt (Abbildung von K. Kuleshov).

Das bedeutet nicht, dass Nanopore Daten nutzlos sind; es gibt nicht bei allen Bakterienspezies solche Probleme. Derzeit scheint in vielen Fällen der beste Workflow

ein Hybridassembly von Nanopore- und Illumina-Sequenzen mit SPAdes zu sein. Nanopore Sequenzdaten wurden genutzt, um z.B. die Gesamtgenome von *Entomospira* Spezies darzustellen, die einem neu beschriebenen Spirochäten Genus angehören (Grana et al. Manuskript in Präparation), oder das Genom einer neuen *Corynebakterium* Spezies (Dangel et al. Manuskript in Präparation). Die in diesem Projekt erhaltenen Daten deuten darauf hin, dass ein Großteil der Probleme mit Nanopore in der Assemblierung der Sequenzdaten liegt (Anhang2\_BakterienHybridAssembly). So wurde z.B. die Entstehung von Chimärsequenzen in Plasmiden bei Verwendung des Canu-asmsemblers festgestellt (Abb. 13). Außerdem wurden mit dieser Software für bestimmte Plasmide mehrere Contigs gebildet, die sich in der Länge unterscheiden (Tabelle 7). Dies stellt kein Problem dar, wenn gute Referenzen zum Vergleich vorliegen, kann jedoch bei der Untersuchung von neuen Spezies oder bei Vorliegen von Rekombinationsereignissen ein schwerwiegendes Problem darstellen. Hier werden derzeit neue Softwarealgorithmen ausprobiert.



**Abbildung 13:** Dot Plot von SPAdes und Canu Assembly. Die Sequenz, die mit Canu assembliert wurde, scheint zweimal in unterschiedlichen Richtungen vorhanden zu sein und könnte auf eine Chimär-Assembly hinweisen.

**Tabelle 7:** Vergleich von *B. burgdorferi* B31 Referenz mit SPAdes und Canu Assemblies von Nanopore Sequenzen. Diese Tabelle verdeutlicht, dass das SPAdes Hybridassembly das Hauptchromosom sowie einige Plasmide sehr gut darstellen kann (z.B. lp54, lp38, cp26), dass aber andere Plasmide nur bruchstückhaft vorhanden sind (z.B. lp17, cp32-5+1). Das Assembly mittels Canu Software produziert zwei unterschiedliche Kategorien, unitigs und contigs. Es sind sehr viel mehr unitigs vorhanden als contigs und es ist zu sehen, dass für einige Plasmide mehrere unitigs vorhanden sind (z.B. lp17, cp26, cp32-5+1). Bei der Erstellung der contigs hat sich zwar die Anzahl der contigs für die meisten Plasmide verringert (z.B. nur noch ein contig für lp32-5+1), für contigs wie cp26 waren aber immer noch 4 contigs vorhanden, die sich in der Größe unterscheiden. Zwei der Plasmide (gelb markiert, nd), die in der untig Fraktion noch vorhanden waren, sind durch die Erstellung von contigs verschwunden, und zwar cp32-2 und cp32-9.

genome element	B31-NRZ [PacBio-Illumina] size [bp]	B31-138 SPAdes hyb	size	size	canu unitigs n=60	size	canu contigs n=20	size
chromosome	909921	1	910017	chromos	tig1, tig2, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 58, 59,60	108000; 51490; 28605; 39313; 53, 22651, 22652, 22653, 22654		
lp56	52951	2	91730	lp36-lp56	tig39, 45,38 (chimär),	37058; 53042; 58132; 29204; 5622649, 22653	lp56	29418; 53071
lp54	53650	3	53733	lp54	tig41	53919	59	53932
lp38	38834	4	39707	lp38	tig36	53019	53	53105
lp36	36780	2	91730	lp36-lp56			58 lp36	89881
lp28-1	21891	?		lp28-1	?		62	51278
lp17	16821	10	10814	lp17	tig24; 23	17676; 29204	32	27308
cp26	26498	7	26625	cp26	tig3, 4, 5, 22, 16, 17; 18; 19; 20;21	35826; 26102; 38749; 32533; 2622, 23, 27, 30		50126, 27728, 26812, 32987
cp32-5-1	61418	8	14957	cp31-5-1	tig31, 29, 26, 32, 33	56541; 30208, 32146 20398; 12	33	66132
cp32-2	31245	12	5513	cp32-2?	tig27, (25 PFam32 cp32-7)	30386, 33372,	nd	
cp32-3	30222	6	28441	cp32-3	tig35	40067	45	40067
cp32-4	30300	9	13850	cp32-4	tig34; (28 - PFam32 cp32-5)	40081; 29828	34	40081
cp32-9	30651	5	30771	cp32-9	tig30	30236	nd	
			11	8795	cp32-5-1?			
			13	3850	cp32-2?			
			14	2575	cp32-4?			
			15					
			16	2240	cp32-2?			

## 4 Zusammenfassung und Schlussfolgerungen

Zusammenfassend ist zu sagen, dass die Nanopore-Sequenzierung erfolgreich am LGL etabliert worden ist. Es ist eine sehr flexible Methode, die es erlaubt, Einzelisolate sowie mehrere Isolate parallel und schnell zu sequenzieren. Es wurden verschiedene Sequenzierungsläufe durchgeführt und erste Erfahrungen in der Handhabung dieser Technologie gesammelt. Die erstellten Daten wurden herangezogen, um methodische Unterschiede in Libraryproduktion zu beurteilen und Vergleiche zur Zweckmäßigkeit verschiedener Assemblierungssoftware anzustellen. Obwohl noch Probleme bei der Assemblierung von Nichtkern-Genombereichen bei Borrelien vorliegen (Anhang2) und hier Verbesserungsbedarf besteht, hat die Assemblierung mit Nanoporesequenzen bei anderen Bakterien recht gut funktioniert (*Entomospira* sp., Corynebakterien) und die Erstellung von vorläufigen Genomen ermöglicht. Diese Ergebnisse werden demnächst in internationale Publikationen einfließen. Die Methode ist auch für die Sequenzierung von MLST PCR Produkten anwendbar; hier werden derzeit noch Erfahrungen mit der Datenauswertung gesammelt.

Eine wichtige Quelle für Probleme liegt in der Aufreinigung der DNA, hier muss unbedingt vermieden werden, während des Aufreinigungsprozesses kleine DNA Fragmente zu produzieren. Dies kann wahrscheinlich mit Hilfe von Pipetten mit weiten Öffnungen vermindert werden. Diesbezüglich werden derzeit noch weitere Versuche unternommen und Methoden weiter ausprobiert.

Proben, die in dieser Studie bisher beispielhaft bearbeitet wurden, waren Borrelien und Corynebakterien. Die an diesen Organismen getesteten Methoden können nun auf andere humanpathogene und ÖDG-relevante Erreger übertragen werden. Daten zur Methodenverbesserung von Vollgenomsequenzierung sind für alle Infektionserreger im Bereich ÖGD von hohem Interesse; NGS wird mit großer Sicherheit mittelfristig im Bereich Public Health/ÖGD eine führende Methodik werden. Es ist daher von Interesse für andere zoonotische oder direkt übertragene Erreger, und auch für Zoonosen, die von anderen Vektoren übertragen werden z.B. Moskitos (z.B. Vogelgrippe, West Nil Virus, ZIKA). Mit unserem Pilotprojekt haben wir einen Kristallisationskern innerhalb des LGL für diese moderne Technologie und ihre Anwendung geschaffen, die früher oder später für praktisch alle Infektionserreger von Bedeutung sein wird. Nachdem das LGL im Bereich Borrelien und Diphtherieerreger aktuell eine international führende Rolle innehat, ist die Übertragung dieses Wissens auch auf andere Erreger möglich und

notwendig. Dies ist auch im Sinne des bayerischen ÖGD als Vorreiterrolle im Bereich Public Health zu sehen und im Hinblick auf die Einrichtung einer NGS Core Unit für Sequenzierung am LGL.

Manuskripte, die aus dieser Arbeit derzeit entstehen:

- 1) Lucia Grana, Silvie Sikutova, Volker Fingerle, Santiago Castillo-Ramirez, Gabriele Margos, Ivo Rudolf: Spirochaetal bacteria isolated from arthropods constitute a novel genus *Entomospira* genus novum within the order Spirochaetales.
- 2) A Dangel, A Berger, J Rau, T Eisenberg, P Kämpfer, G Margos, R Konrad, M Peters, R Sting, M Contzen, A Sing: A unique group of NTTB *Corynebacterium* spp. in wild boar and roe deer suggests delineation of the a new species, *Corynebacterium silvaticum* sp. nov.

Danksagung

Dank geht an Alexander Vogel, Rheinisch-Westfälische Technische Hochschule Aachen, und Konstantin Kuleshov, Central Research Institute of Epidemiology, Moskau, Russland, für Hilfe bei der bioinformatischen Auswertung von Sequenzdaten.

## 5 Literatur

- Bankevich, A., S. Nurk, et al. (2012). "SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing." *J Comput Biol* **19**(5): 455-477.
- Becker, N. S., G. Margos, et al. (2016). "Recurrent evolution of host and vector association in bacteria of the *Borrelia burgdorferi* sensu lato species complex." *BMC Genomics* **17**(1): 734.
- Casjens, S. R., E. F. Mongodin, et al. (2012). "Genome stability of Lyme disease spirochetes: comparative genomics of *Borrelia burgdorferi* plasmids." *PLoS ONE* **7**(3): e33280.
- Castillo-Ramirez, S., V. Fingerle, et al. (2016). "Trans-Atlantic exchanges have shaped the population structure of the Lyme disease agent *Borrelia burgdorferi* sensu stricto." *Sci Rep* **6**: 22794.
- de Lannoy, C., D. de Ridder, et al. (2017). "The long reads ahead: *de novo* genome assembly using the MinION." *F1000Res* **6**: 1083.
- Fraser, C. M., S. Casjens, et al. (1997). "Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*." *Nature* **390**(6660): 580-586.
- Heather, J. M. and B. Chain (2016). "The sequence of sequencers: The history of sequencing DNA." *Genomics* **107**(1): 1-8.



- Hodzic, J., L. Gurbeta, et al. (2017). "Overview of Next-generation Sequencing Platforms Used in Published Draft Plant Genomes in Light of Genotypization of Immortelle Plant (*Helichrysum Arenarium*)."  
Med Arch **71**(4): 288-292.
- Jackson, R. W., B. Vinatzer, et al. (2011). "The influence of the accessory genome on bacterial pathogen evolution."  
Mob Genet Elements **1**(1): 55-65.
- Jolley, K. A. (2009). "Internet-based sequence-typing databases for bacterial molecular epidemiology."  
Methods Mol Biol **551**: 305-312.
- Jolley, K. A. and M. C. Maiden (2010). "BIGSdb: Scalable analysis of bacterial genome variation at the population level."  
BMC Bioinformatics **11**: 595.
- Kolmogorov, M., J. Yuan, et al. (2019). "Assembly of long, error-prone reads using repeat graphs."  
Nat Biotechnol.
- Koren, S., B. P. Walenz, et al. (2016). "Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation."  
bioRxiv.
- Levy, S. E. and R. M. Myers (2016). "Advancements in Next-Generation Sequencing."  
Annu Rev Genomics Hum Genet **17**: 95-115.
- Margos, G., S. Hepner, et al. (2017). "Lost in plasmids: next generation sequencing and the complex genome of the tick-borne pathogen *Borrelia burgdorferi*."  
BMC Genomics **18**(1): 422.
- Marosevic, D., G. Margos, et al. (2017). "First insights in the variability of *Borrelia recurrentis* genomes."  
PLoS Negl Trop Dis **11**(9): e0005865.
- Moon, B. Y., J. Y. Park, et al. (2016). "Mobilization of Genomic Islands of *Staphylococcus aureus* by Temperate Bacteriophage."  
PloS ONE **11**(3): e0151409.
- Quail, M. A., M. Smith, et al. (2012). "A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers."  
BMC Genomics **13**: 341.
- Reuter, J. A., D. V. Spacek, et al. (2015). "High-throughput sequencing technologies."  
Mol Cell **58**(4): 586-597.
- Rhoads, A. and K. F. Au (2015). "PacBio Sequencing and Its Applications."  
Genomics Proteomics Bioinformatics **13**(5): 278-289.
- Strube, C., V. M. Montenegro, et al. (2010). "Establishment of a minor groove binder-probe based quantitative real time PCR to detect *Borrelia burgdorferi sensu lato* and differentiation of *Borrelia spielmanii* by ospA-specific conventional PCR."  
Parasit Vectors **3**: 69.
- Tyler, S., S. Tyson, et al. (2018). "Whole genome sequencing and phylogenetic analysis of strains of the agent of Lyme disease *Borrelia burgdorferi* from Canadian emergence zones."  
Sci Rep **8**(1): 10552.