

Aus dem Lehrstuhl für
Medizinische Mikrobiologie und Hygiene
der Fakultät für Medizin
der Universität Regensburg
Direktor: Prof. Dr. Dr. André Gessner

Entwicklung eines *Next-Generation-Sequencing*-Protokolls für die Genotypisierung
von Hepatitis-A-Viren und Hepatitis-E-Viren unter Verwendung der *Nanopore*-
Technologie

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Zahnmedizin

der Fakultät
für Medizin
der Universität Regensburg

vorgelegt von
Annika Renner

2026

Aus dem Lehrstuhl für
Medizinische Mikrobiologie und Hygiene
der Fakultät für Medizin
der Universität Regensburg
Direktor: Prof. Dr. Dr. André Gessner

Entwicklung eines *Next-Generation-Sequencing*-Protokolls für die Genotypisierung
von Hepatitis-A-Viren und Hepatitis-E-Viren unter Verwendung der *Nanopore*-
Technologie

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Zahnmedizin

der Fakultät
für Medizin
der Universität Regensburg

vorgelegt von
Annika Renner

2026

Dekan: Prof. Dr. Dirk Hellwig

1. Berichterstatter: Prof. Dr. Jürgen Wenzel

2. Berichterstatter: PD Dr. Florian Hitzenbichler

Tag der mündlichen Prüfung: 13.08.2026

Inhalt

1	Einleitung	5
1.1	Das Hepatitis-A-Virus	5
1.2	Das Hepatitis-E-Virus	7
1.3	Next-Generation-Sequencing	10
2	Material und Methoden	13
2.1	Material	13
2.1.1	Verbrauchsmaterialien	13
2.1.2	Kits und Reagenzien	13
2.1.3	Technische Geräte	14
2.1.4	Programme	15
2.1.5	Klinisches Probenmaterial (HAV und HEV)	16
2.1.6	Sanger-Sequenzdaten	16
2.2	Methoden	17
2.2.1	Verwendete DNA-Fragmente zur Genotypisierung: HAV net, HEV net, VP1 und ORF1	17
2.2.2	DNA-Konzentrationsbestimmung (Qubit-4-System)	18
2.2.3	Rapid Barcoding und NGS mittels ONT	18
2.2.4	Auswertung der Nanopore-Sequenzdaten	19
2.2.5	Percent Identity Matrix	27
2.2.6	Phylogenetische Analyse	27
2.2.7	Methodik der Workflow-Erstellung und Textbearbeitung	27
3	Ergebnisse	29
3.1	Bewertung der automatisierten Auswertungspipeline	29
3.2	Einfluss der Probenanzahl pro Sequenzierungslauf auf die medianen Lesetiefen	29
3.3	Vergleich der Lesetiefen zwischen unaufgereinigten Amplifikaten und Macherey-Nagel-Aufreinigungen	31
3.4	Einfluss der DNA-Konzentration auf die Lesetiefen	32
3.5	Verlauf der Lesetiefe über das HAV VP1- und net-Fragment	33
3.6	Zusammenhang zwischen der Referenzsequenz, der medianen Lesetiefe und der Percent Identity	34

3.7	Validierung negativer Proben sowie Betrachtung der Anzahl der N-Nukleotide und der durchschnittlichen Lesetiefe als Qualitätskriterien	36
4	Diskussion	41
5	Zusammenfassung	50
6	Summary	51
7	Anhang	52
a.	Abbildungsverzeichnis	52
b.	Phylogenetische Analysen	53
c.	Abkürzungsverzeichnis	57
8	Literaturverzeichnis	58
9	Danksagung	68
10	Eidesstattliche Erklärung	69

1 Einleitung

1.1 Das Hepatitis-A-Virus

Das Hepatitis-A-Virus (HAV) ist weltweit eine der häufigsten Ursachen für eine akute Hepatitis-Infektion (1). Es kommt in Ländern aller sozioökonomischen Entwicklungsstufen vor. Obwohl Verbesserungen im Bereich Hygiene, der sicheren Trinkwasserversorgung und die Verfügbarkeit von effektiven Impfstoffen zu einem Rückgang in den Industrieländern geführt haben, stellt das Virus in vielen Entwicklungsländern weiterhin eine Herausforderung dar (2). Klinische Beschreibungen der Hepatitis A lassen sich bis in die Antike zurückverfolgen (3). Der direkte Virusnachweis gelang jedoch erst 1973 durch Feinstone et al. mittels Immunelektronenmikroskopie (4). Die WHO klassifiziert das globale Infektionsrisiko anhand einer dreistufigen Endemieskala in hohe, mittlere und niedrige Endemizität (1,5). Hochendemiegebiete für Hepatitis-A-Infektionen liegen vor allem in Subsahara-Afrika sowie in Teilen Südasiens wie Indien, Pakistan oder Bangladesch. In diesen Regionen liegt die Seroprävalenz innerhalb der ersten fünf Lebensjahre nahezu bei 100 % (6). Regionen wie Nordafrika, Mittel- und Südamerika sowie Osteuropa weisen hingegen ein mittleres Risiko für Hepatitis A auf (7).

Während jüngere Kinder meist symptomfrei bleiben, nehmen die Symptome mit zunehmendem Alter an Schwere zu und reichen von grippeähnlichen Symptomen wie Fieber, Abgeschlagenheit, Appetitlosigkeit, Übelkeit, Erbrechen und Bauchschmerzen bis hin zu akutem Leberversagen (1). In Europa hat Hepatitis A eine relativ niedrige Letalität zwischen 0,03–0,26 % (8), wohingegen in aktuelleren Ausbrüchen in den USA eine Letalität von bis zu 1 % berichtet wurde (9). Weltweit treten Ausbrüche vor allem dort auf, wo der Zugang zu sauberem Trinkwasser und angemessener Hygiene eingeschränkt ist. Die Übertragung erfolgt auf fäkal-oralem Weg, meist über kontaminiertes Wasser oder Lebensmittel wie Meeresfrüchte oder Tiefkühlbeeren (1). Darüber hinaus ist eine Übertragung durch direkten Kontakt möglich, beispielsweise im gemeinsamen Haushalt (10), in Gemeinschafts- und Pflegeeinrichtungen sowie durch sexuelle Kontakte (11). Ergänzend wurde in Einzelfällen auch über HAV-

Übertragungen durch Blutprodukte berichtet (12).

HAV (*Hepatitis A virus*) ist ein quasi-behülltes (13), 27–32 nm großes RNA-Virus aus der Gattung *Hepatitis A virus* der Familie *Picornaviridae* (14). HAV-Partikel zeichnen sich durch eine hohe Stabilität gegenüber Umwelteinflüssen wie Raumtemperatur und einem niedrigen pH-Wert aus, was die Übertragung zusätzlich begünstigt (15). Das positive einzelsträngige RNA-Genom ist ungefähr 7,5 kb lang. Das Genom gliedert sich in eine 5'-UTR, einen zentralen offenen Leserahmen mit den Regionen P1–P3 und einer 3'-UTR mit Poly-A-Schwanz (siehe Abbildung 1) (3).

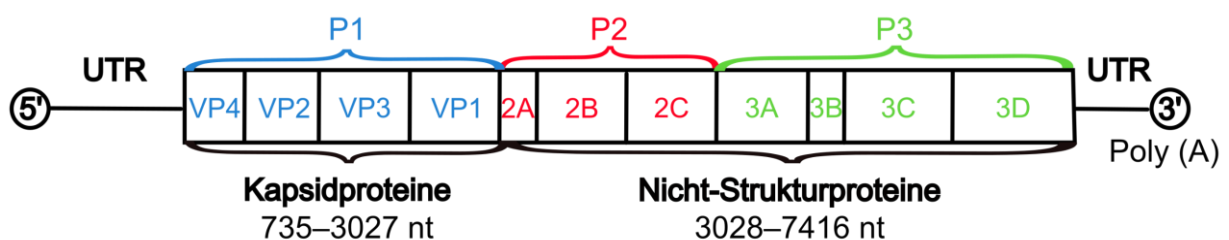


Abbildung 1: Schematische Darstellung des HAV-Genoms, adaptiert von Hussain et al., 2011 (16).

Das ~7,5 kb lange HAV-Genom beginnt mit einer 5'-UTR, gefolgt von einem offenen Leserahmen und endet mit einer 3'-UTR. Der offene Leserahmen kodiert für Kapsidproteine (VP1–4) und Nicht-Strukturproteine (P2A–3D).

UTR: untranslated region; nt: Nukleotide

Eine Besonderheit von HAV ist das Vorkommen in zwei Formen: einerseits unbehüllte Virionen im Stuhl infizierter Personen sowie quasi-umhüllte Virionen im Blut, Urin oder in Zellkultur-Überständen (14). Es existiert zwar nur ein Serotyp, jedoch sechs Genotypen. Dabei sind nur die Genotypen I, II und III humanpathogen, wobei diese zusätzlich jeweils in die Subtypen IA, IB, IIA, IIB, IIIA und IIIB unterteilt werden (siehe Abbildung 2) (14,17). Genotyp IV stammt aus einem Javaneraffen (*Macaca fascicularis*) von den Philippinen, wohingegen Genotyp V aus einer grünen Meerkatze (*Cercopithecus aethiops*) aus Kenia isoliert wurde. Genotyp VI stammt ebenfalls aus einem Javaneraffen (*Macaca fascicularis*), der ursprünglich jedoch aus Indonesien kommt (18). Die präzise Identifizierung dieser Genotypen ist für das Verständnis des Infektionsgeschehens entscheidend, stellt jedoch hohe Anforderungen an die molekulare Diagnostik.

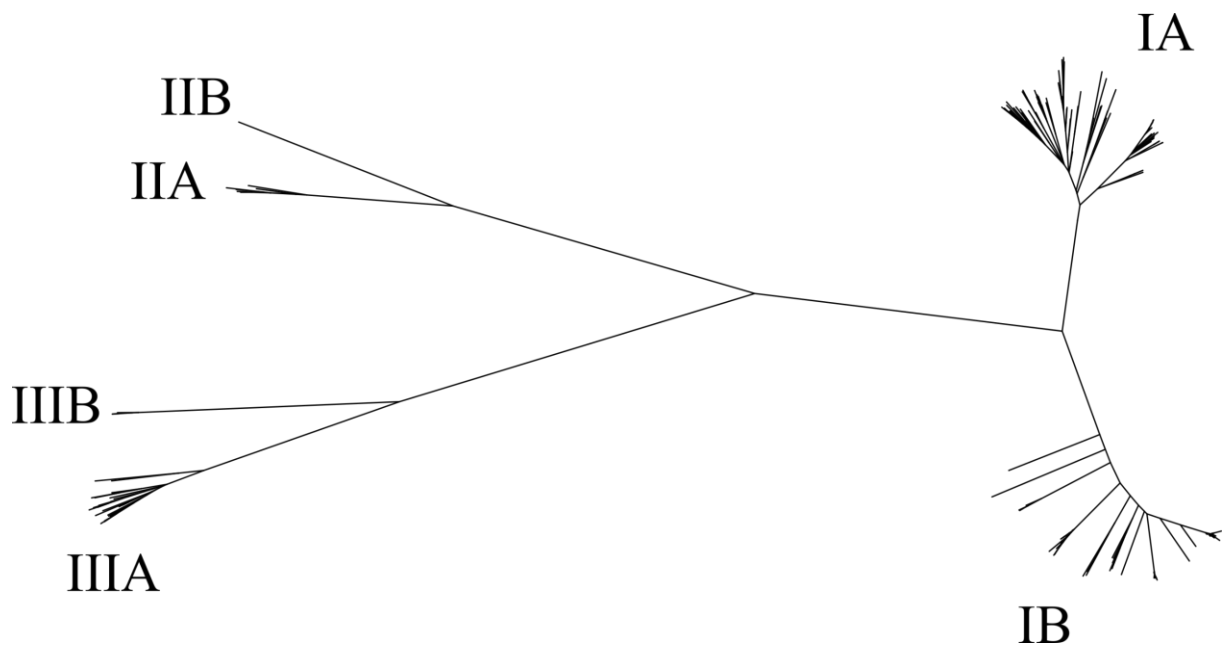


Abbildung 2: Phylogenetische Darstellung der humanpathogenen Genotypen I – III.
 Eigene Darstellung in Anlehnung an (14,17).

1.2 Das Hepatitis-E-Virus

Das Hepatitis-E-Virus (HEV) ist weltweit die Hauptursache einer akuten Hepatitis (19,20). Das Virus gehört zur Familie *Hepeviridae* und ist als *Paslahepevirus balayani* klassifiziert (21). Durch einen wassergebundenen Ikterus-Ausbruch in Indien, bei dem serologische Tests weder Erreger für Hepatitis A noch Hepatitis B als Ursache nachweisen konnten, wurde 1978 das HEV, damals als Non-A- und Non-B-Hepatitis bezeichnet, erstmals postuliert (22). 1983 konnte Balayan schließlich den fäkal-oralen Übertragungsweg bestätigen (23), während erste retrospektive Berichte auf 1955 datiert wurden (20,24). HEV verursacht in Entwicklungsregionen regelmäßig große Ausbrüche, von denen mehrere Hundert bis Tausend Menschen betroffen sein können. Ursache dafür ist eine fäkal-orale Übertragung durch kontaminiertes Wasser, insbesondere während der Regenzeit, wenn Trinkwasserquellen durch Überschwemmungen fäkal verunreinigt werden (25). Im Gegensatz dazu wurde HEV in Industrieländern lange Zeit primär als „Reisekrankheit“ betrachtet, gilt jedoch inzwischen auch dort als endemisch (26). Allerdings erfolgt die Übertragung fast

ausschließlich alimentär über unzureichend gegarte Haus- oder Wildschweinprodukte. Trotz der inzwischen verpflichtenden Testung von Blutprodukten (Blutspende, Organspende) auf HEV kann es in Einzelfällen weiterhin zu parenteralen Übertragungen kommen (27). Seit 2001 unterliegen Infektionen mit HEV in Deutschland einer Meldepflicht. Nach einem kontinuierlichen Anstieg der gemeldeten HEV-Fälle bis 2019 auf 3.728 Fälle pro Jahr kam es in den Folgejahren 2020 bis 2022 zunächst zu einem leichten Rückgang auf Werte zwischen 3.081 und 3.514 Fällen pro Jahr. Ab 2023 verzeichnete das Robert Koch-Institut jedoch einen erneuten deutlichen Anstieg, der 2024 mit 4.754 gemeldeten Infektionen ein Maximum erreichte. Dieser Trend setzt sich im Jahr 2025 fort. Für das erste Halbjahr 2025 wurden bislang 2.989 Fälle registriert, was einer deutlichen Steigerung gegenüber dem Vergleichszeitraum 2019 mit 1.927 Fällen entspricht. (Robert Koch-Institut: SurvStat@RKI 2.0, abrufbar unter <https://survstat.rki.de/Content/Query/Create.aspx> , Abfragedatum: 13.08.2025).

Schätzungen zufolge lebt etwa ein Drittel der Weltbevölkerung in Endemiegebieten und ist somit einem Infektionsrisiko ausgesetzt (19). Ein besonders hohes Risiko für schwere Verläufe besteht in bestimmten Endemiegebieten für schwangere Frauen, bei denen die Letalität signifikant erhöht ist (28,29).

HEV ist ein kleines, quasi-behülltes (30), circa 27–34 nm großes Virus und besitzt ein einzelsträngiges (+)-RNA-Genom von etwa 7,2 kb Länge. Das Genom umfasst eine kurze 5'-UTR mit einer 7-Methylguanosin-Kappe, drei ORFs sowie eine kurze 3'-UTR, die mit einem Poly-A-Schwanz abschließt (31). Der ORF1 kodiert für ein nicht-strukturelles Polyprotein, ORF2 für das Kapsidprotein und ORF3 für ein multifunktionelles Phosphoprotein (siehe Abbildung 3) (32).

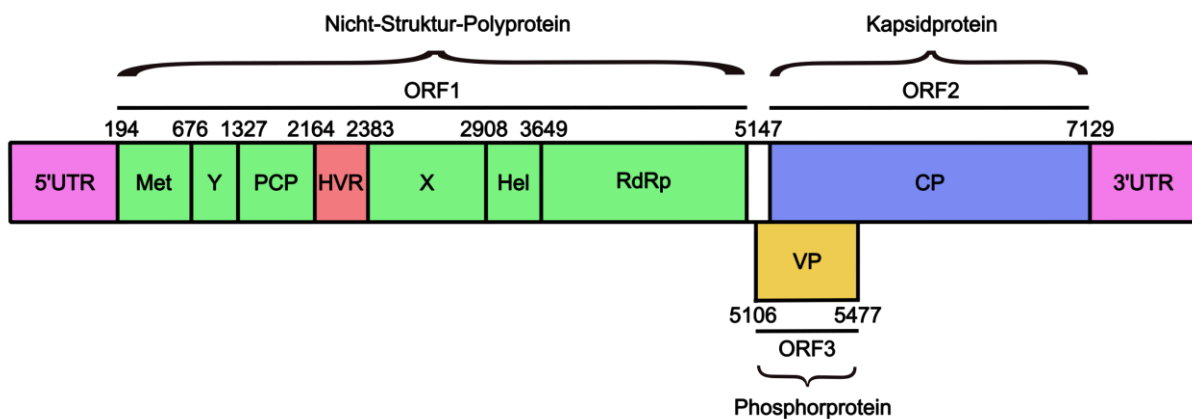


Abbildung 3: Schematische Darstellung des HEV-Genoms, adaptiert von Laugel et al., 2021 (33).

Gezeigt sind die funktionellen Domänen anhand des HEV-Genoms mit der GenBank-Nummer M73218 (Burma, Genotyp 1). ORF1 kodiert für ein Nicht-Struktur-Polypeptid mit den Domänen Methyltransferase (Met), Y-Domäne (Y), Papain-ähnliche Cystein-Protease (PCP), hypervariable Region (HVR), X-Domäne (X), Helikase (Hel) und RNA-abhängige RNA-Polymerase (RdRp). ORF2 kodiert für das Kapsidprotein (CP). ORF3 kodiert für ein multifunktionelles Phosphoprotein (VP).

Insgesamt werden acht verschiedene Genotypen (HEV-1 bis -8) sowie 36 Subtypen unterschieden. Während üblicherweise nur HEV-1 bis -4 als humanpathogen gelten (34), wurde mit HEV-7 bis jetzt nur ein Mensch nachweislich infiziert, wobei es sich um einen Patienten nach Lebertransplantation mit regelmäßigem Konsum von Kamelfleisch und -milch handelte (35). HEV-1 und -2 sind vor allem in Entwicklungsregionen verbreitet, wo sie fäkal-oral primär über kontaminiertes Wasser von Mensch zu Mensch übertragen werden und zu großen Ausbrüchen führen können (24,36). Im Gegensatz dazu werden HEV-3 (weltweit) und HEV-4 (vornehmlich in Asien) zoonotisch übertragen. Die Übertragung geschieht hier entweder durch direkten Kontakt mit infizierten Tieren oder durch den Verzehr kontaminierter, unzureichend gegarter Schweinefleischprodukte (37,38). Als weitere Reservoirwirte, in denen HEV-3 und -4 identifiziert wurden, dienen unter anderem Wildschweine, Ziegen, Rinder oder Hirsche, wobei sich das bekannte Wirtsspektrum kontinuierlich erweitert (38,39). In Deutschland ist HEV-3 der prävalente Genotyp, wobei die Subtypen in absteigender Häufigkeit HEV-3c, -3f, -3e und -3i dominieren (40).

1.3 Next-Generation-Sequencing

Der oben beschriebene Anstieg der Infektionszahlen, vor allem bei HEV, erfordert eine kontinuierliche und effiziente molekulare Überwachung, da Viruspanidemien die Menschheitsgeschichte seit jeher begleiten. Durch die Entschlüsselung genetischer Informationen lassen sich Aufbau und Funktionsweise von Viren heute präzise verstehen. Dies erlaubt nicht nur Rückschlüsse auf Übertragungswege und Replikationsmechanismen, sondern bildet auch die essenzielle Grundlage für zielgerichtete Diagnostik, Therapie und Impfstoffentwicklung (41,42). Besonders vor dem Hintergrund zunehmender Reisetätigkeit und Globalisierung ist dieses molekulare Monitoring unverzichtbar geworden, um auf neu auftretende Bedrohungen reagieren zu können (41).

Während die Sequenzierung des ersten menschlichen Genoms über ein Jahrzehnt dauerte und Kosten von fast drei Milliarden US-Dollar verursachte (43), ist eine Sequenzierung heute innerhalb weniger Tage zu deutlich geringeren Kosten möglich. Maßgeblich hierfür ist die stetige Weiterentwicklung des sogenannten *Next-Generation-Sequencing* (NGS) (44,45). Diese Technologie ermöglicht Millionen paralleler Sequenzierungsreaktionen und führt somit zu einer signifikanten Effizienzsteigerung gegenüber klassischen Methoden (46).

Als erste Generation wird die Sanger-Sequenzierung bezeichnet, die aufgrund der präzisen Didesoxymethode bis heute als Goldstandard vor allem für gezielte Einzelsequenzierungen gilt (44,47). Das Verfahren ist jedoch pro Ansatz auf eine einzige Sequenzierungsreaktion begrenzt. Das Prinzip beruht auf der Verwendung spezifischer Terminator-Nukleotide (Didesoxynukleotide), denen die notwendige 3'-OH-Gruppe für die Ausbildung einer Phosphodiesterbindung fehlt. Der Einbau dieser Nukleotide durch die DNA-Polymerase führt zum Abbruch der Kettenverlängerung (42). Um die resultierenden Fragmente unterschiedlicher Länge anschließend zu analysieren, werden sie heute mit fluoreszierenden Farbstoffen markiert und mittels Kapillarelektrophorese separiert und detektiert (48).

Die Methoden der zweiten Generation werden auch als Short-Read-Technologien bezeichnet (42,46). Der derzeitige Marktführer ist die Firma Illumina, welche eine

Sequenzierung durch Synthese mit optischer Detektion verwendet. Dieses Verfahren erinnert konzeptionell an die Sanger-Sequenzierung, weist jedoch einen deutlich höheren Durchsatz auf (46). Diese Short-Read-Technologien zeichnen sich durch eine hohe Genauigkeit auf Basenebene aus, sind jedoch durch die geringen Leselängen begrenzt. Die kurzen Reads müssen nachträglich wieder zusammengesetzt werden, weshalb ihre Anwendung insbesondere bei hochkomplexen strukturellen Variationen des Genoms an ihre Grenzen stößt (42).

Im Gegensatz dazu können die Long-Read-Technologien der dritten Generation, wie jene von Pacific Biosciences (PacBio) und Oxford Nanopore Technologies (ONT), deutlich längere Sequenzabschnitte in einem Lauf erfassen. Dies bietet insbesondere bei der Analyse komplexerer Genomstrukturen signifikante Vorteile (49). Die Komplexität der Datenzusammenstellung von Short-Read-Sequenzierungen, insbesondere bei großen genomischen Umlagerungen, kann durch die Long-Read-Sequenzierung überwunden werden (46,50). Wenngleich diese Technologien derzeit noch eine höhere Fehlerrate aufweisen, ermöglichen Systeme wie die SMRT-Technologie (PacBio) oder der MinION (ONT) eine Einzelmolekül-Echtzeitsequenzierung mit signifikant gesteigerten Leselängen (42,51). Das MinION-System nutzt hierfür ein Gitter aus in Membranen eingebetteten biologischen Nanoporen, an denen eine elektrische Spannung anliegt (46,51). Nach dem Einbringen der Probe führt der Durchtritt der einzelsträngigen DNA-Fragmente zu einer für die jeweilige Basenabfolge charakteristischen Änderung des Ionenstroms, die anschließend mit einer Software bioinformatisch ausgewertet und in eine Sequenz übersetzt werden kann (sogenanntes *Basecalling*) (49,50).

Gerade angesichts der globalen Mobilität ist eine schnelle, zuverlässige und kostengünstige Diagnostik sowie eine kontinuierliche molekulare Überwachung von Hepatitis-A- und Hepatitis-E-Viren zum Schutz der öffentlichen Gesundheit von großer Bedeutung (1,20). Moderne Sequenzierverfahren bieten hierfür eine entscheidende technologische Grundlage (41). Zwar gilt die Sanger-Sequenzierung als sehr zuverlässiges Verfahren, sie erfordert jedoch einen hohen manuellen Laboraufwand sowie einen beträchtlichen personellen Ressourcenaufwand. Aufgrund der schwierigen Automatisierbarkeit ist sie für die großflächige Routineanwendung

weniger geeignet. Im Gegensatz dazu lässt sich die Auswertung von NGS-Daten, wenngleich erst nach aufwendiger Etablierung automatisierter bioinformatischer Auswertungsprozesse (sogenannte *Pipelines*), weitgehend vollautomatisch durchführen (42). Dabei muss allerdings bei NGS-Verfahren häufig eine Mindestprobenmenge erreicht werden, um den Prozess kosteneffizient zu gestalten, was zu einer zeitlichen Verzögerung führen kann (52).

Ziel der vorliegenden Arbeit war es daher, ein qualitativ gleichwertiges Sequenzierverfahren für HAV- und HEV-Fragmente als Alternative zur Sanger-Sequenzierung zu etablieren. Dieses neue Verfahren soll durch einen hohen Automatisierungsgrad mit deutlich weniger manuellem Arbeitsaufwand einhergehen und so die Effizienz in der Diagnostik steigern.

2 Material und Methoden

2.1 Material

2.1.1 Verbrauchsmaterialien

MATERIAL	HERSTELLER
Handschuhe	
Gentle Skin® grip	Meditrade, Kiefersfelden, Deutschland
Peha-soft® nitrile fino	Paul Hartmann, Heidenheim, Deutschland
Pipettenspitzen	
SafeSeal Tips Professional 10 µl	Biozym Scientific, Hessisch Oldendorf, Deutschland
SafeSeal Tips Professional 100 µl	Biozym Scientific, Hessisch Oldendorf, Deutschland
SafeSeal Tips Professional 300 µl	Biozym Scientific, Hessisch Oldendorf, Deutschland
SafeSeal-Tips Professional 1250 µl	Biozym Scientific, Hessisch Oldendorf, Deutschland
Röhrchen	
8er PCR Kette	Sarstedt AG, Nümbrecht, Deutschland
1,5 ml DNA LoBind® Tubes	Eppendorf, Hamburg, Deutschland

2.1.2 Kits und Reagenzien

BEZEICHNUNG	HERSTELLER
Next Generation Sequencing	
Rapid Barcoding Kit 96 V14 (SQK-RBK114.96)	Oxford Nanopore Technologies, Oxford, Großbritannien
Qubit 1X dsDNA HS Working Solution	Thermo Scientific, Waltham, Massachusetts, USA
Qubit 1X dsDNA BR Working Solution	Thermo Scientific, Waltham, Massachusetts, USA
Ultrapure BSA non-acetylated	Thermo Scientific, Waltham, Massachusetts, USA
Flow Cells	
MinION GridION Flow Cell R10 Version Pk.8	Oxford Nanopore Technologies, Oxford, Großbritannien

Sonstige	
EMSURE ACS ISO Ethanol 99 %	Merck KGaA, Darmstadt, Deutschland
Nuklease-freies Wasser	Ambion, Austin, Texas, USA

2.1.3 Technische Geräte

GERÄT	HERSTELLER
Zentrifugen	
Eppendorf Zentrifuge 5417 C	Eppendorf, Hamburg, Deutschland
Eppendorf Zentrifuge 5430 R	Eppendorf, Hamburg, Deutschland
MiniFuge Plus	Starlab, Hamburg, Deutschland
Mini-Zentrifuge Sprout	Biozym, Hessisch Oldendorf, Deutschland
Thermocycler	
PEQSTAR 2X Gradient	VWR International, Ismaning, Deutschland
Schüttler	
Vortex Genie 2	Scientific Industries, Bohemia, New York, USA
ZX3 Vortex Mixer	VELP Scientifica Srl, Usmate Velate, Italien
HulaMixer Probenmischer	Thermo Fisher Scientific, Waltham, Massachusetts, USA
DNA-Konzentrationsbestimmung	
Qubit 4 Fluorometer	Thermo Fisher Scientific, Waltham, Massachusetts, USA
Pipetten	
Eppendorf Reference 2 10 µl	Eppendorf, Hamburg, Deutschland
Eppendorf Reference 2 20 µl	Eppendorf, Hamburg, Deutschland
Eppendorf Reference 2 100 µl	Eppendorf, Hamburg, Deutschland
Eppendorf Reference 2 200 µl	Eppendorf, Hamburg, Deutschland
Eppendorf Reference 2 1000 µl	Eppendorf, Hamburg, Deutschland
Eppendorf Research plus 0,5–10 µl	Eppendorf, Hamburg, Deutschland
Magnetischer Ständer	
CleanMag Magnetic Plate	Paragon Genomics, Brandin Court Fremont, Kalifornien, USA
Sequenziergerät	

MinION Mk1C	Oxford Nanopore Technologies, Oxford, Großbritannien
Sonstiges	
Elements Festplatte 1 TB extern	Western Digital, San José, Kalifornien, USA
Eismaschine	Ziegra, Isernhagen, Deutschland
Kühlschrank	Liebherr, Bulle, Schweiz
Ultratiefkühlschrank Innova U725	Eppendorf, Hamburg, Deutschland
High-End Workstation	Mifcom, München, Deutschland
EliteDesk 800 G5 SFF PC	HP, Palo Alto, Kalifornien, USA
Dell LCD P2213 Desktop	Dell Technologies, Round Rock, Texas, USA

2.1.4 Programme

SOFTWARE	VERSION	ERSTELLER
AliView	1.28	Anders Larsson (53)
Citavi	6.17	Swiss Academic Software
Cowplot	1.1.3	Claus Wilke https://wilkelab.org/cowplot/
Dorado	0.5.3	Oxford Nanopore Technologies
Dplyr	1.1.4	Wickham et al. https://cran.r-project.org/web/packages/dplyr/index.html
Fasta36	36.3.8h	William R. Pearson (54)
FigTree	1.4.4	Andrew Rambaut http://tree.bio.ed.ac.uk/software/figtree/
R	4.1.2	R Core Team (55) https://www.r-project.org/
Ggplot2	3.5.1	Hadley Wickham (56)
Inkscape	0.92	Inkscape Community
Ivar	1.4.2	Grubaugh et al. (57)
LibreOffice	7.3	Document Foundation, Berlin, Deutschland
MAFFT	7	Kazutaka Katoh (58,59)
Minimap2	2.26	Heng Li (60,61)
Office	2016	Microsoft Corporation, Redmond, Washington, USA
Porechop	0.2.4	Bonenfant, Quentin, Noé Laurent; Touzet Hélène (62)

Python	3.9.7	Van Rossum, G. & Drake Jr, F. L., Amsterdam, Niederlande https://www.python.org/
RAxML	8.2.11	Alexandros Stamatakis (63)
RAxML-ng	1.2.1	Alexey M Kozlov (64)
Samtools	1.20	Heng Li (65)
SigmaPlot	13.0	Systat Software Inc., Palo Alto, Kalifornien, USA

2.1.5 Klinisches Probenmaterial (HAV und HEV)

Die pseudonymisierten Amplifikate für HAV und HEV aus den klinischen Untersuchungsmaterialien wurden von der Virologie des Instituts für klinische Mikrobiologie und Hygiene am Universitätsklinikum Regensburg (IMHR) zur Verfügung gestellt. Dabei handelte es sich um Restmaterial aus der Diagnostik. Bis zur Verwendung wurden die Proben kurzfristig im Kühlschrank bei 7 °C oder langfristig bei –20 °C gelagert.

2.1.6 Sanger-Sequenzdaten

Als Referenz für die Etablierung des neuen Verfahrens dienten die Sanger-Sequenzdaten, die am IMHR aus denselben Proben generiert und bereitgestellt wurden. Für die Proben erfolgte zunächst eine Nukleinsäureisolation, woraufhin eine Amplifikation der Zielregionen und eine Aufreinigung mittels der Kits von Macherey-Nagel folgten. Anschließend wurde die Sequenz mittels Sanger-Sequenzierung im Rahmen der Diagnostik am IMHR ausgewertet. Diese Arbeitsschritte waren nicht Bestandteil der experimentellen Durchführung der vorliegenden Arbeit.

2.2 Methoden

2.2.1 Verwendete DNA-Fragmente zur Genotypisierung: HAV net, HEV net, VP1 und ORF1

Zur Identifikation der unterschiedlichen Genotypen der amplifizierten HAV- und HEV-positiven Proben wurden spezifische genomische Zielregionen untersucht. Die Sequenzierung charakteristischer Abschnitte ermöglicht sowohl die phylogenetische Einordnung als auch Vergleiche der Proben untereinander (66).

Zudem wurden in die Analyse auch negative Proben miteinbezogen. Diese dienen nicht der Genotypisierung, sondern zur Qualitätssicherung als Kontrollen, um die Spezifität des Verfahrens zu überprüfen und potenzielle Kontaminationen innerhalb des Workflows weitgehend auszuschließen.

Bei HEV-Proben wurden das 242 bp lange ORF1-Fragment (67) sowie das 493 bp lange HEV net-Fragment (68) analysiert (siehe Abbildung 4 A).

Bei HAV-Proben wurden das 460 bp lange HAV net-Fragment (40,66) sowie das mit den Nukleotiden 1 bis 349 des HAV net-Amplikons überlappende VP1-Fragment untersucht (siehe Abbildung 4 B).

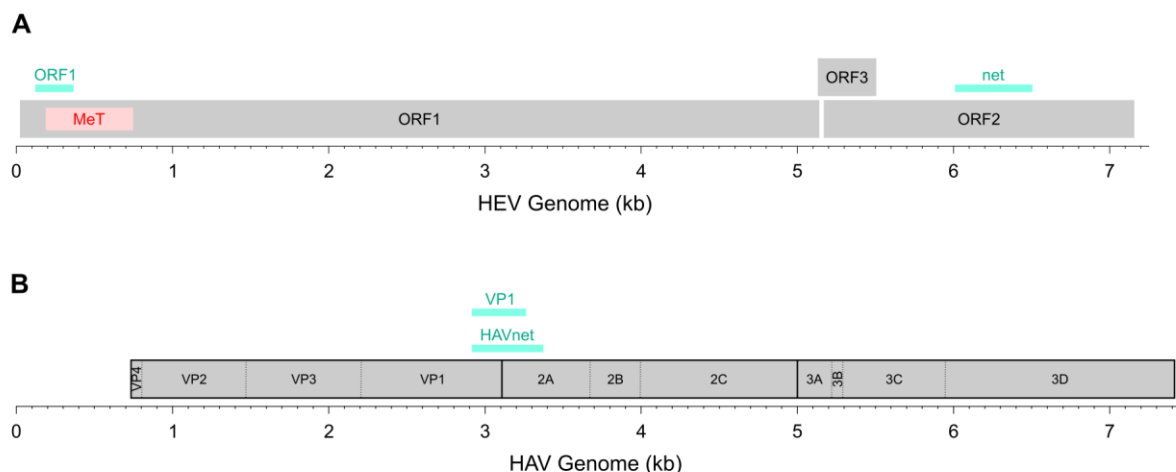


Abbildung 4: Schematische Darstellung der Genomorganisation von HEV und HAV.

(A) HEV-Genom, modifiziert nach Schemmerer et al. 2022 (40). Lokalisation der amplifizierten Fragmente ORF1 und HEV net in Bezug zu den ORFs auf dem HEV-Genom.

(B) HAV-Genom, eigene Darstellung. Lokalisation des amplifizierten HAV net-Fragments und des überlappenden VP1-Fragments auf dem HAV-Genom.

HEV: Hepatitis-E-Virus; HAV: Hepatitis-A-Virus; kb: Kilobasen; MeT: Methyltransferase; ORF: open reading frame

2.2.2 DNA-Konzentrationsbestimmung (Qubit-4-System)

Die DNA-Konzentration der Amplikons wurde mithilfe des Qubit-4-Fluorometers bestimmt. Je nach erwarteter Konzentration kamen dabei die Qubit 1X dsDNA HS (High Sensitivity) oder BR (Broad Range) Assay-Kits zum Einsatz.

Nach dem einmaligen Kalibrieren mit zwei herstellerepezifischen Standards zu Beginn wurden pro zu messender Probe 199 µl der Qubit 1X dsDNA HS Working Solution in einem beschrifteten Qubit-Tube vorgelegt. Anschließend wurde jeweils 1 µl der Probe hinzu pipettiert. Die Ansätze wurden danach für fünf Sekunden durch Vortexen gemischt und kurz abzentrifugiert. Nach einer Inkubationszeit von zwei Minuten bei Raumtemperatur (RT) erfolgte die Messung der DNA-Konzentration im Qubit-4-Fluorometer. Bei Proben mit einer sehr hohen Konzentration (bzw. unverdünnten Proben) wurde statt der HS Working Solution die BR Working Solution entsprechend verwendet.

2.2.3 Rapid Barcoding und NGS mittels ONT

Die Multiplex-Sequenzierung erfolgte mittels *Nanopore*-Technologie auf einem *MinION Mk1C* unter Verwendung des *Rapid Barcoding Kits 96 V14 (SQK-RBK114.96)*. Zur Gewährleistung einer gleichmäßigen Abdeckung wurden die Amplikons initial auf eine einheitliche DNA-Konzentration eingestellt, wobei dieser Schritt im siebten Durchlauf bewusst weggelassen wurde. Im Falle von HEV wurden die jeweiligen PCR-Produkte (ORF1 und HEV net) vorab zu gleichen Verhältnissen gepoolt.

Für die Barcodierung wurden die Proben mit den zugehörigen Rapid Barcodes in einem Thermocycler (2 Minuten bei 30 °C, 2 Minuten bei 80 °C, 1 Minute bei 4 °C) inkubiert. Zur anschließenden Aufreinigung der barcodierten Ansätze wurden AMPure XP Beads im Verhältnis 1:1 dazu gemischt, gefolgt von einer zweifachen Waschung mit 80%igem Ethanol und einer Resuspension in Elution-Buffer. Für die finale Ligation wurden vom Eluat maximal 800 ng DNA des Rapid Adapters eingesetzt und mit dem Sequencing Buffer und den Library Beads zusammengestellt.

Nach erfolgreichem Flow Cell Check (> 800 aktive Poren) und dem zweistufigen

Priming der Flow Cell mit Flow Cell Flush, Bovine Serum Albumin (BSA) und Flow Cell Tether wurde die Library über den SpotON Sample Port geladen. Das BSA dient zur Blockierung unspezifischer Bindungsstellen und damit der Steigerung der Ausbeute. In der Software MinKNOW wurden die Laufparameter wie Dauer, Speicherort und -format festgelegt sowie das automatische *Basecalling* deaktiviert, um die Rohdaten für die spätere bioinformatische Auswertung zu sichern (69).

2.2.4 Auswertung der Nanopore-Sequenzdaten

Das Ziel der vorliegenden Doktorarbeit war die Automatisierung der Auswertung für NGS-Daten von HAV- und HEV-Proben sowie deren Subtyp-Identifizierung und Konsensus-Generierung. Hierfür wurde ein parametrisiertes Skript erstellt, das durch Platzhalter an unterschiedliche Eingabedateien angepasst werden kann. Dabei wird die Zuordnung der Probennamen zu den verwendeten Barcode-Nummern und dem zu sequenzierenden Virus (HAV oder HEV) über eine zentrale Konfigurationsdatei (im vorliegenden Fall „barcodes.csv“) gesteuert.

Um die hohe Sequenzvariabilität von HAV und HEV abzudecken, erfolgte das Mapping nicht gegen eine Einzelsequenz, sondern gegen ein repräsentatives Set aus Referenzsequenzen. Diese sind in den folgenden Tabellen aufgeführt. Für HEV wurden 44 humanpathogene Genotypen nach Smith et al. (2020) verwendet (34). Die HAV-Referenzsequenzen entsprechen den Subtyp-Definitionen des HAV-Net (Global Hepatitis A Virus Laboratory-based Surveillance Network) (70) und repräsentativen Isolaten aus der GenBank.

HAV-Subtyp	Isolat	Accession
HAV-IA	AH2	AB020565.1
HAV-IA	FG	X83302.1
HAV-IA	FH2	AB020568.1
HAV-IA	HAV5	EU131373.1
HAV-IB	HM-175	M14707.1
HAV-IB	IVA	DQ646426.1
HAV-IB	LA1	AF314208.1
HAV-IIA	CF53-Berne	AY644676.1
HAV-IIB	SLF88	AY644670.1
HAV-IIIA	GBS-IND	DQ991030.1
HAV-IIIA	HA-JNG08-92	AB279733.1
HAV-IIIA	NOR-21	AJ299464.3
HAV-IIIB	HA-JNG06-90F	AB258387.1

HAV-IIIB	HAI85-1	AB279735.1
HAV-IIIB	KRM003G72	AB425339.1
HAV-V	AGM-27	D00924.1
HAV-V	IND-SHAV	EU140838.1

HEV-Subtyp	Isolat	Accession
HEV-1a	Burma	M73218.1
HEV-1b	Xinjiang	NC_L08816.1
HEV-1c	I1	X98292.1
HEV-1d	Morocco	AY230202.1
HEV-1e	T3	AY204877.1
HEV-1f	IND-HEV-AVH5-2010	JF443721.1
HEV-1g	MNE15-072	LC225387.1
HEV-1uc1	HEV-H	FJ457024.1
HEV-1uce	NG/17-0503	MH918640.1
HEV-2a	Mex-14	KX578717.1
HEV-2b	NG/17-0500	MH809516.1
HEV-3a	Meng	AF082843.1
HEV-3b	JRA1	AP003430.1
HEV-3c	wbGER27	FJ705359.1
HEV-3e	swJ8-5	AB248521.1
HEV-3f1a	E116-YKH98C	AB369687.1
HEV-3f1b	SW627	EU723513.1
HEV-3f2a	swX07-E1	EU360977.1
HEV-3f2b	FR-R	KJ873911.1
HEV-3g	Osh205	AF455784.1
HEV-3h1	TR19	JQ013794.1
HEV-3i	BB02	FJ998008.1
HEV-3j	Arkell	AY115488.1
HEV-3k	E088-STM04C	AB369689.1
HEV-3l	FR-SHEV3c-like	JQ953664.1
HEV-3m	IC2011	KU513561.1
HEV-3ra	GDC9	FJ906895.1
HEV-3uc1	swMN06-C1056	AB290313.1
HEV-3uc2	MWP-2010	KP294371.1
HEV-3uc3	WB/HEV/NA17ITA15	MF959764.1
HEV-3uc4	WB/HEV/NA21ITA15	MF959765.1
HEV-3uc5	swHE1606845	LC260517.1
HEV-3uc6	17RS1920	MK390971.1
HEV-4a	JKO-ChiSai98C	AB197673.1
HEV-4b	swDQ	DQ279091.2
HEV-4c	JAK-Sai	AB074915.3
HEV-4d	T1	AJ272108.1
HEV-4e	IND-SW-00-01	AY723745.1
HEV-4f	HE-JA2	AB220974.1
HEV-4g	CCC220	AB108537.1
HEV-4h	CHN-XJ-SW13	GU119961.3
HEV-4i	E067-SIJ05C	AB369690.1
HEV-4uc1a	E087-SAP04C	AB369688.1
HEV-4uc1b	SWU/42-5/2018	MK410048.1

Im Folgenden werden nun die einzelnen Schritte der erstellten bioinformatischen Auswertung dargestellt. Abbildung 5 zeigt einen Überblick über diese nacheinander ablaufenden Schritte: Nach Abschluss aller Vorbereitungen endet das Bash-Skript mit dem Mapping an die oben aufgeführten Referenzsequenzen unterschiedlicher Subtypen. Daraufhin werden die Lesetiefen bestimmt und mit dem R-Skript der Subtyp mit der höchsten medianen Lesetiefe ermittelt. Die Konsensus-Sequenz des Subtyps mit der höchsten medianen Lesetiefe wird dann durch das Python-Skript in einer finalen Datei ausgegeben.

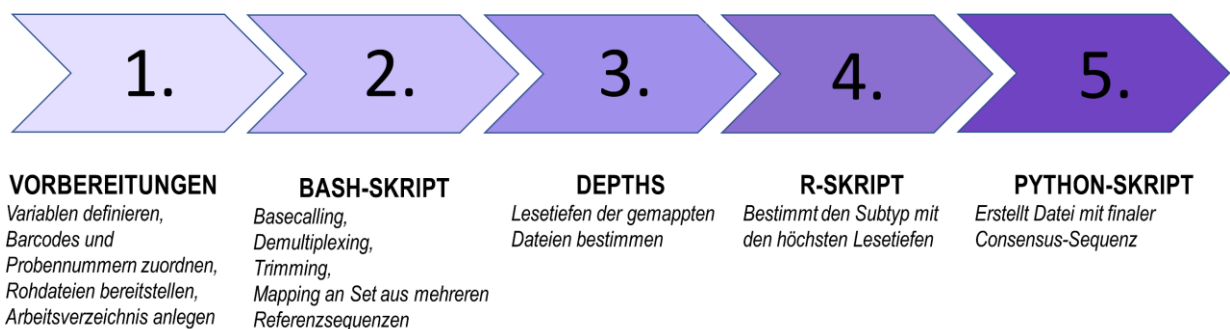


Abbildung 5: Schematischer Ablauf der bioinformatischen Auswertung.

Der Workflow startet mit den Vorbereitungen, woraufhin das Basecalling, Demultiplexing, Trimming und das Mapping gegen ein repräsentatives Set aus Referenzsequenzen folgt (Bash-Skript). Nach der Bestimmung der Lesetiefen (Depths) wird automatisiert der Subtyp mit der höchsten Abdeckung identifiziert (R-Skript) und dessen finale Konsensussequenz in einer Ergebnisdatei ausgegeben (Python-Skript).

Abbildung 8 visualisiert die spezifischen Befehlsabfolgen sowie die methodischen Unterschiede bei der automatisierten Auswertung von HAV- und HEV-Proben von den Rohdateien der Nanoporesequenzierung bis zur finalen Konsensussequenz:

Die Auswertung startete mit dem *Duplex Basecalling* der POD5-Rohdateien, gefolgt vom Demultiplexing mittels *dorado* von ONT. Dabei wurden die Sequenzdaten der einzelnen Proben aus dem Multiplex-Pool extrahiert. Dieses Programm war in der Arbeitsgruppe bereits etabliert. Nach einer Formatumwandlung wurden die Adapter- und Primersequenzen mit der Software *porechop* entfernt. Das Trimmen dieser Sequenzen ist in nahezu allen NGS-basierten Analysen ein grundlegender und unerlässlicher Schritt, um Artefakte in der nachfolgenden Konsensusbildung zu vermeiden.

Es erfolgte das Mapping gegen ein Set aus repräsentativen Referenzsequenzen verschiedener HAV- und HEV-Subtypen mit *minimap2*. Aufgrund der erheblichen

Sequenzvariabilität beider Viren ist die Nutzung eines breiten Referenzspektrums unerlässlich, um die vielfältigen genetischen Varianten und Subtypen bei der Analyse abzudecken.

Im nächsten Schritt wurden mittels *samtools view* die nicht gemappten Sequenzen entfernt, bevor die verbleibenden gemappten Reads mittels *samtools sort* koordinatenbasiert sortiert wurden.

Ab diesem Punkt verzweigte sich die Prozessierung je nach Virusart:

Bei HEV-positiven Proben wurde mit *samtools view* der notwendige Header in die Dateien integriert, um die korrekte Referenzzuordnung zu gewährleisten.

Von den Dateien einer HAV-positiven Probe wurde zuerst der Primer mittels *ivar trim* entfernt. Nach einer erneuten Dateikonvertierung mit *samtools view* mussten die HAV-Dateien noch einmal mit *samtools sort* sortiert werden.

Anschließend wurde aus diesen aufbereiteten Ausgabedateien beider Viren mit *samtools consensus* für jeden Subtyp eine Konsensussequenz generiert. Im Rahmen der Methodenoptimierung wurden initial auch alternative Programme wie *sam2consensus.py* und *ivar consensus* evaluiert, wobei *samtools consensus* die stabilsten Ergebnisse lieferte.

Zur finalen Bestimmung der Subtyp-Zugehörigkeit der Proben wurde im nächsten Schritt die Lesetiefe (Depth) der gemappten Fragmente ermittelt. Dies ist ein weiterer wichtiger Qualitätsparameter, da die Reads im Mapping-Prozess aufgrund hoher Sequenzähnlichkeiten oft an mehrere mögliche Referenzsequenzen unterschiedlicher Subtypen mappen, wodurch für jede Referenzsequenz eine eigene Konsensussequenz generiert wird. Der Vergleich der Lesetiefen dient als entscheidendes Qualitätskriterium für die finale Subtyp-Zuordnung und die Validierung der Sequenzgenauigkeit. Die Lesetiefe gibt an, wie oft die einzelnen Nukleotide des Fragments durch Sequenzierungsreads abgedeckt sind. Somit ist sie ein Maß für die Sequenziergenauigkeit und -verlässlichkeit und stellt einen zentralen Parameter vor allem in der Hochdurchsatzsequenzierung dar (71).

Die Lesetiefe kann über das gesamte Fragment mittels *samtools depth* ermittelt werden. Diejenige Referenzsequenz, die über alle Nukleotidpositionen hinweg die höchste Lesetiefe aufwies, wurde als diejenige des korrekten Subtyps der unbekannten Probe definiert.

Für die Identifizierung des korrekten Subtyps wurden anschließend die medianen und durchschnittlichen Lesetiefen zu jedem Subtyp einer untersuchten Probe berechnet. Um diesen Schritt zu automatisieren, wurde ein R-Skript entwickelt, das die Lesetiefen entlang der Referenzsequenz analysiert und den Subtyp mit der höchsten medianen und durchschnittlichen Lesetiefe identifiziert. Die Nutzung des Medians dient als robuster statistischer Parameter, um den Einfluss lokaler Sequenzierungsartefakte auf die Subtyp-Zuordnung zu minimieren. Ziel dieses R-Skriptes war es, die Grundlage für die Extraktion der finalen Konsensussequenz aus der wahrscheinlichsten Subtypzuordnung zu schaffen.

Skripte in R werden vor allem für Datenanalysen, Visualisierungen und statistische Berechnungen verwendet (72). Das entwickelte Skript importierte die tabellarischen Rohdaten der Lesetiefen zu den verschiedenen Subtypen und verarbeitete diese in mehreren Schritten: Die Lesetiefe wurde über die gesamte Fragmentlänge hinweg für jede Referenzsequenz berechnet und mittels *dplyr* (Gruppierung und Berechnung der Daten), *ggplot2* (Erstellung des Plots) und *cowplot* (Veränderung des Layouts des Plots) grafisch dargestellt (siehe Abbildung 6). Um eine objektive Subtyp-Identifizierung zu ermöglichen, generierte das Skript danach eine zusammenfassende Statistikdatei. In dieser Datei sind zu jeder Referenzsequenz die mediane, durchschnittliche sowie die maximale und minimale Lesetiefe aufgeführt, wobei sie für eine bessere Vergleichbarkeit nach absteigendem Median sortiert ist (siehe Abbildung 7). Diese Ergebnisse wurden automatisiert als PDF-Grafiken und im TSV-Format gespeichert. Zusätzlich wurden für den Überblick alle Statistik-Dateien aus einem Lauf in einer Datei konsolidiert, um eine barcode-sortierte Gesamtübersicht des Sequenzierlaufs zu erstellen.

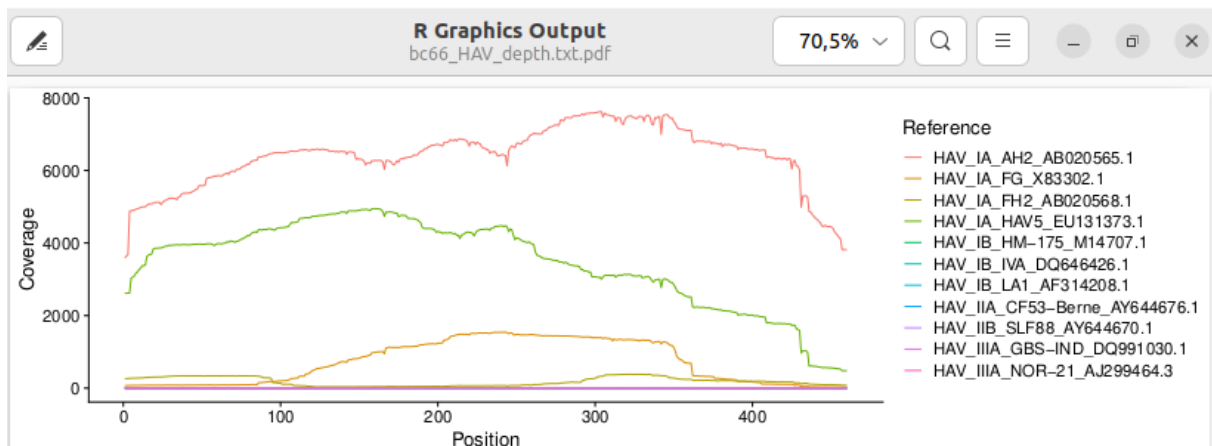


Abbildung 6: Plot von Barcode 66 (ONT 7).

Lesetiefenwerte eines gemapten HAV-net-Fragments (460 Basenpaare) als Beispiel zur Veranschaulichung gezeigt. Man erkennt, dass die Referenzsequenz des Wildtyp IA_AH2 (rot) deutlich die höchsten Lesetiefen über die gesamte Länge des Fragments aufweist.

bc: Barcode; HAV: Hepatitis-A-Virus; ONT: Oxford Nanopore Technologies

id	Median	Mean	Max	Min
<fct>	<dbl>	<dbl>	<dbl>	<dbl>
1 HAV_IA_AH2_AB020565.1	6530.	6393.	7622	3589
2 HAV_IA_HAV5_EU131373.1	3939	3467.	4948	478
3 HAV_IA_FG_X83302.1	612.	706.	1544	30
4 HAV_IA_FH2_AB020568.1	166.	180.	381	32
5 HAV_IB_HM-175_M14707.1	3	2.89	5	1
6 HAV_IB_IVA_DQ646426.1	3	2.71	5	0
7 HAV_IIIA_NOR-21_AJ299464.3	2	1.79	3	1
8 HAV_IIB_SLF88_AY644670.1	1	1.36	3	0
9 HAV_IIIA_GBS-IND_DQ991030.1	0.5	0.86	3	0
10 HAV_IB_LA1_AF314208.1	0	0.5	1	0
11 HAV_IIA_CF53-Berne_AY644676.1	0	0.27	1	0

Abbildung 7: Tabelle mit ausgewerteten Referenzsequenzen von Barcode 66 (ONT 7) (Statistikdatei).

Ansicht im Terminal, nachdem der Befehl „map.R“ ausgeführt wurde. Unter „id“ sind alle Referenzsequenzen aufgeführt, gegen die gemappt wurde. Danach folgt die mediane Coverage („Median“) und die durchschnittliche Lesetiefe („Mean“) sowie der Maximalwert („Max“) und der Minimalwert („Min“) über die gesamte gemappte Referenzsequenz hinweg. Man erkennt, wie auch schon in Abbildung 6 dargestellt, dass der Wildtyp IA_AH2 die höchste Coverage aufweist.

id: Identifikation; ONT: Oxford Nanopore Technologies

Basierend auf der vorangegangenen statistischen Bewertung wurde anschließend die finale Konsensussequenz mit einem eigens erstellten Python-Skript extrahiert. Ziel dieses Skriptes war es, ausschließlich die Sequenz des passenden Subtyps mit der höchsten medianen Lesetiefe in einer separaten Ergebnisdatei auszugeben.

Das Skript verknüpfte die Information aus der Statistik-Datei mit den zuvor generierten Konsensussequenzen aller Subtypen. Dabei identifizierte es die Referenzsequenz mit der höchsten medianen Lesetiefe und extrahierte die zugehörige Konsensussequenz

isoliert in eine finale Ergebnisdatei. Dadurch wird eine eindeutige, qualitativ hochwertige Referenzsequenz für weitere phylogenetische und vergleichende Analysen bereitgestellt. Zur Implementierung wurden Funktionen der Python-Standardbibliothek für die Verzeichnisnavigation und reguläre Ausdrücke für die automatisierte Mustererkennung genutzt, um eine fehlerfreie Zuordnung über große Datensätze hinweg zu gewährleisten.

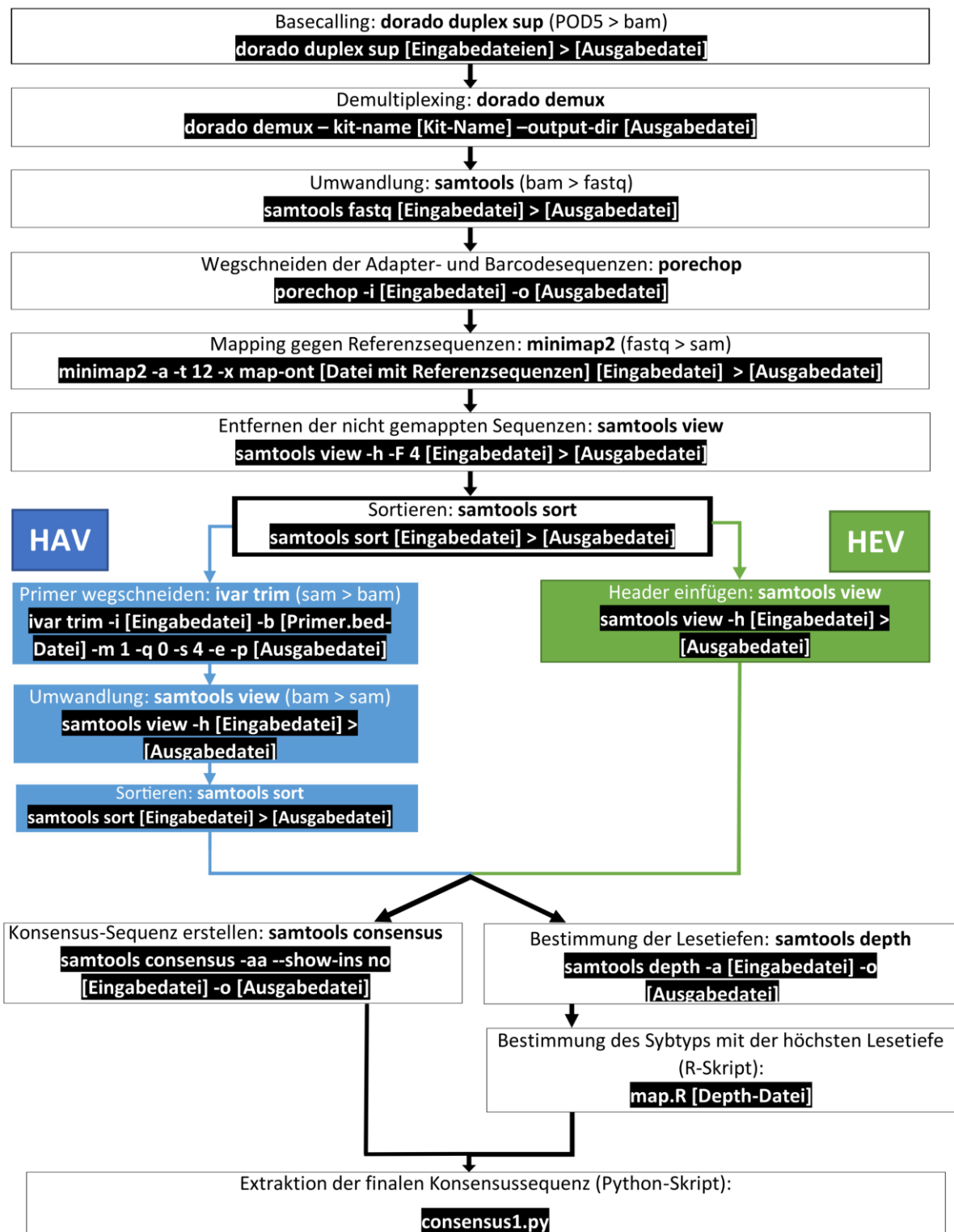


Abbildung 8: Übersicht Workflow bioinformatische Auswertung von den Roh-POD5-Dateien bis zur Konsensussequenz.

2.2.5 Percent Identity Matrix

Zur Validierung der generierten Konsensussequenzen mit der eigens erstellten bioinformatischen Auswertung wurde ein Vergleich mit den bereits verfügbaren Sanger-Sequenzen durchgeführt. Dazu wurden die Konsensussequenzen über ein MUSCLE-Alignment (73) aneinander ausgerichtet und mithilfe von *clustalo* die prozentuale Übereinstimmung wie folgt berechnet.

```
clustalo --infile -force [Eingabedatei] --verbose --percent-id --distmat-  
out [Ausgabedatei als Matrix]--full --resno --output-order input-order -  
outfile [Ausgabedatei als Clustal] --seqtype dna -force
```

2.2.6 Phylogenetische Analyse

Für die phylogenetische Rekonstruktion wurden die Gesamtgenomsequenzen mittels *MAFFT* aligniert. Anschließend wurden die Alignments mit *MAFFT --addfragments* erweitert, indem die Sanger-Fragment-Sequenzen aus der Diagnostik und die NGS-Fragmente dieser Doktorarbeit integriert wurden. In diesem Prozess wurden auch formal negative Proben miteinbezogen, um eine umfassende Vergleichsbasis zu schaffen. Das resultierende Alignment wurde danach auf die relevante Fragmentgröße beschnitten. Der phylogenetische Baum wurde anschließend unter Verwendung von *RAXML-NG* berechnet.

2.2.7 Methodik der Workflow-Erstellung und Textbearbeitung

Die Entwicklung des bioinformatischen Workflows erfolgte unter Nutzung verschiedener digitaler Ressourcen. Zur Einarbeitung in den Stand der Technik, die Erstellung, Optimierung und Fehlerbehebung wurden öffentlich zugängliche Repositories (z. B. GitHub), Software-Dokumentationen (Manuals) und Fachforen herangezogen. Die Erstellung und Optimierung der Skripte (Bash, R, Python) sowie die abschließende sprachliche Überarbeitung der schriftlichen Ausarbeitung (Rechtschreibprüfung, Syntaxoptimierung und stilistisches Lektorat) erfolgten unter

Zuhilfenahme der generativen KI-Plattformen ChatGPT sowie Gemini. Die damit generierten Code-Bausteine und Textvorschläge wurden stets manuell geprüft, modifiziert, an die lokale Infrastruktur sowie die fachlichen Inhalte angepasst und durch systematische Testläufe mit Referenzdatensätzen validiert. Die finale inhaltliche Kontrolle, die Dateninterpretation sowie die Verantwortung für die Richtigkeit der Ergebnisse verblieben zu jedem Zeitpunkt vollständig bei der Autorin.

3 Ergebnisse

3.1 Bewertung der automatisierten Auswertungspipeline

Der zuvor entwickelte, vollständig automatisierte Workflow zur bioinformatischen Auswertung der HAV- und HEV-Sequenzierdaten ermöglichte die erfolgreiche Prozessierung von 262 Sequenzen aus fünf Sequenzierläufen (Läufe 1, 2, 3, 4 und 7) mithilfe eines modular aufgebauten Systems aus Bash-, R- und Python-Komponenten ohne manuelle Eingriffe. Die Läufe 5 und 6 lieferten aufgrund fehlerhafter Flow-Cell-Beladung keine auswertbaren Daten und wurden deshalb nicht berücksichtigt. Die Anwendung standardisierter Dateiformate (z. B. barcodes.csv) und konsistenter Benennungskonventionen ermöglichte eine reibungslose und wiederholbare Auswertung. Der erstellte Workflow stellt damit einen zentralen methodischen Bestandteil dieser Arbeit dar.

3.2 Einfluss der Probenanzahl pro Sequenzierungslauf auf die medianen Lesetiefen

Die Beziehung zwischen der medianen Lesetiefe (y-Achse) und der Probenanzahl (x-Achse) wurde für das HAV net-, HEV net- und HEV ORF1-Fragment in drei Sequenzierläufen mit unterschiedlicher Probenanzahl ($n = 16, 32$ und 48) untersucht und in Abbildung 9 dargestellt. Dabei nahmen die absoluten Lesetiefen mit steigender Probenanzahl generell ab.

Dies wird ebenfalls deutlich in der medianen Lesetiefe aller Proben pro Sequenzierdurchgang (schwarz hervorgehoben): Beim HEV net-Fragment sank sie von durchschnittlich 12.257 ($n = 16$) auf 6.201 ($n = 32$) und schließlich auf 2.985 ($n = 48$). Ähnliche Abnahmen bei steigender Probenanzahl zeigten sich für das HEV ORF1-Fragment: 5.377 bei $n = 16$, 3.632 bei $n = 32$ und 1.412 bei $n = 48$. Auch für HAV sank die Lesetiefe von 10.815 ($n = 16$) über 2.730 ($n = 32$) auf 2.164 ($n = 48$). Obwohl der allgemeine Abwärtstrend in allen Fällen zu beobachten war, unterscheiden sich die absoluten medianen Lesetiefenwerte sowie das Ausmaß der Abnahme

deutlich. So lag die mediane Lesetiefe für HEV net bei $n = 16$ mit 12.257 deutlich über HEV ORF1 mit 5.377, während HAV mit 10.815 einen mittleren Wert einnahm. Besonders auffällig war die relative Abnahme: Während die Lesetiefe bei HAV von $n = 16$ (10.815) auf $n = 32$ (2.730) um 74,8 % sank, verzeichnete das HEV ORF1-Fragment im selben Vergleich lediglich einen Rückgang um 32,5 % (5.377 auf 3.632). Während der Rückgang der Lesetiefe bei steigender Probenanzahl ein erwartetes Phänomen der Multiplex-Sequenzierung darstellt, deuten die variierenden Abnahmeraten darauf hin, dass nicht nur die reine Probenanzahl, sondern auch die Fragment-spezifischen Faktoren (z. B. Primer-Affinität, GC-Gehalt oder Amplikonlänge) die Sequenzierleistung maßgeblich beeinflussen können.

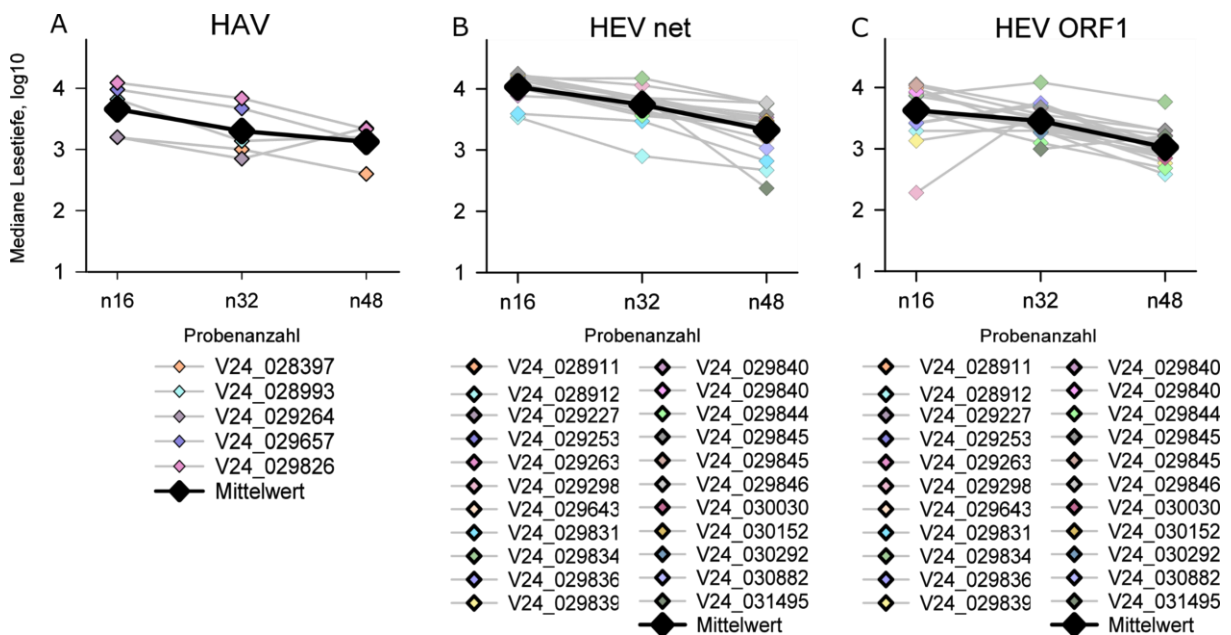


Abbildung 9: Mediane Lesetiefen in Abhängigkeit von der Probenanzahl.

Für jedes Fragment ist ein Diagramm dargestellt. Die y-Achse zeigt die medianen Lesetiefen in logarithmischer Skalierung pro Sequenzierlauf. Auf der x-Achse wird die unterschiedliche Probenanzahl pro Sequenzierlauf beschrieben.

log: dekadischer Logarithmus; n: Anzahl

3.3 Vergleich der Lesetiefen zwischen unaufgereinigten Amplifikaten und Macherey-Nagel-Aufreinigungen

PCR-Produkte können für nachfolgende Amplifikationen aufgereinigt werden. Eine Möglichkeit ist die Aufreinigung mit einem Macherey-Nagel-Kit. Hierfür wurden vier HAV-Proben untersucht, die sowohl im unaufgereinigten Zustand als auch nach einer Aufreinigung mit dem Macherey-Nagel-Kit amplifiziert und sequenziert wurden (Abbildung 10).

Bei drei der vier Proben erzielte die Aufreinigung deutlich höhere mediane Lesetiefen als die jeweiligen unaufgereinigten Ansätze. Die vierte Probe wies hingegen keinen nennenswerten Unterschied zwischen den beiden Varianten auf. Auffällig war hierbei, dass deren mediane Lesetiefe bereits im Ausgangszustand deutlich über dem Niveau der drei anderen Proben lag. Diese Ergebnisse deuten darauf hin, dass eine Aufreinigung insbesondere bei Proben mit geringer Ausgangsqualität die Lesetiefe maßgeblich verbessern kann, bei Proben mit bereits hoher Ausgangsqualität jedoch keinen zusätzlichen Nutzen bietet.

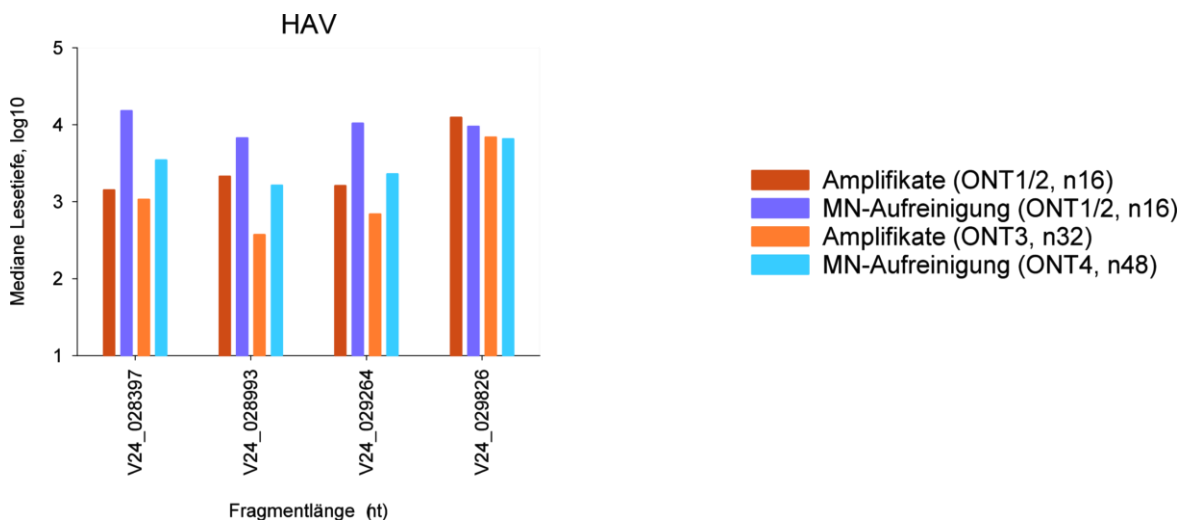


Abbildung 10: Vergleich der medianen Lesetiefen zwischen unaufgereinigten Amplifikaten und Macherey-Nagel-Aufreinigungen.

Die x-Achse zeigt die vier (mehrmals) untersuchten Proben mit pseudonymisiertem Namen, für die jeweils auf der y-Achse die medianen Lesetiefen in logarithmischer Skalierung dargestellt sind. Dabei beschreiben die roten und orangefarbenen Balken die unaufgereinigten Amplifikate, während die lila und hellblauen Balken die entsprechenden aufgereinigten Proben darstellen, wobei jede Probe in vier unterschiedlichen Läufen verwendet wurde

log: dekadischer Logarithmus; MN: Macherey-Nagel; n: Anzahl; nt: Nukleotide; ONT: Oxford Nanopore Technologies

3.4 Einfluss der DNA-Konzentration auf die Lesetiefen

Die Lesetiefen (y-Achse) von HAV-Proben wurden in zwei Gruppen über das gesamte Fragment untersucht: Proben ohne vorherige Verdünnung (Abbildung 11 A) und Proben mit einer Verdünnung auf die von ONT empfohlene DNA-Konzentration (Abbildung 11 B). Im Sequenzierlauf mit den verdünnten Proben wurde pro Probe im Median 4,3 μ l eingesetzt, im Vergleich zu 9 μ l pro Probe bei den unverdünnten Proben. Die verdünnten Proben erzielten insgesamt deutlich geringere Lesetiefen als die unverdünnten Proben. Dies spricht dafür, dass eine höhere eingesetzte DNA-Menge bei gleichbleibender Qualität zu besseren Sequenzierungsergebnissen führen kann, auch wenn die Empfehlung des Herstellers dabei überschritten wird.

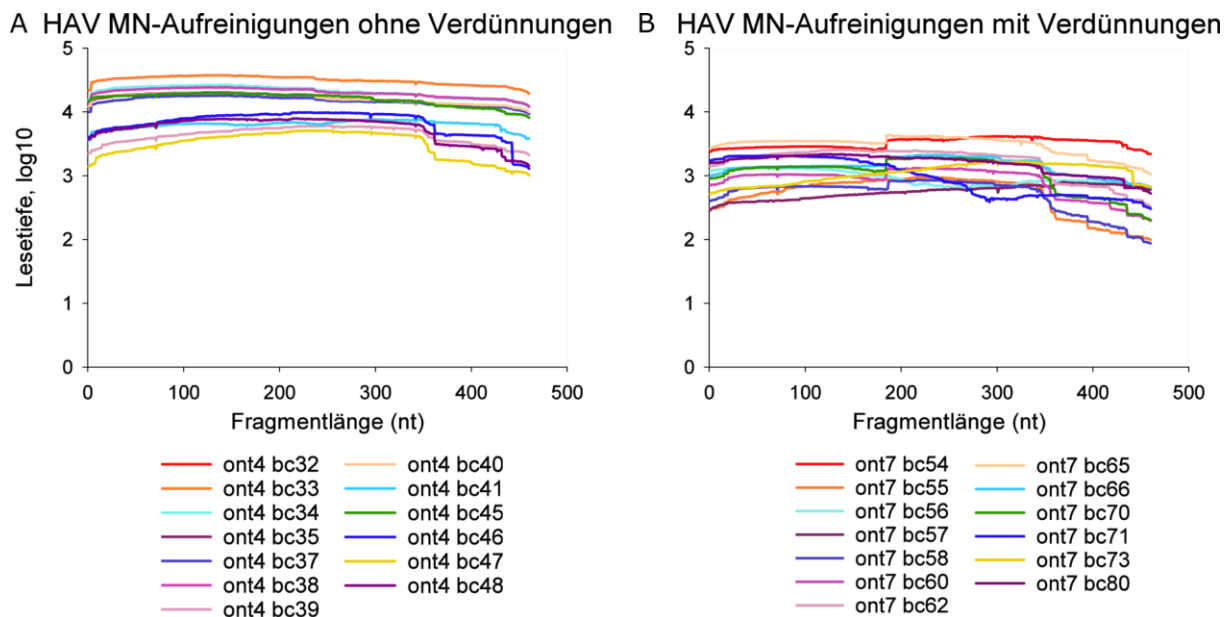


Abbildung 11: Vergleich zwischen unverdünnten und verdünnten Proben.

Es sind zwei Diagramme gegenübergestellt: In A (links) wurden die Proben unverdünnt verwendet, während in B (rechts) die Proben auf die von ONT empfohlene DNA-Konzentration verdünnt wurden. Die x-Achse zeigt die Fragmentlänge in Nukleotiden an. Auf der y-Achse ist die logarithmierte Lesetiefe dargestellt.

bc: Barcode; log: dekadischer Logarithmus; MN: Macherey-Nagel; ont: Oxford Nanopore Technologies; nt: Nukleotide

3.5 Verlauf der Lesetiefe über das HAV VP1- und net-Fragment

Abbildung 12 zeigt den Verlauf der Lesetiefe pro Basenposition über die gesamte Länge der HAV-Fragmente für verschiedene Proben. Dabei ist zu beachten, dass das VP1-Fragment innerhalb des net-Fragments liegt (vergleiche Abbildung 4 B). Bei Proben, die sowohl VP1- als auch net-positiv waren (A), zeigte sich ein markanter Einbruch der Lesetiefe bei etwa 349 Nukleotiden. Dieser Abfall ist durch das Ende des kürzeren VP1-Fragments bedingt. Ein ähnliches Muster trat auch bei Proben auf, die lediglich VP1-positiv waren (B). Auch hier sank die Lesetiefe nach dem 349. Nukleotid deutlich ab. Eine Ausnahme stellte die Probe ONT7 bc92 dar, bei der die Lesetiefe über die gesamte Fragmentlänge weitestgehend konstant blieb.

Im Gegensatz dazu zeigten Proben, die ausschließlich für das net-Fragment positiv und somit für das VP1-Fragment negativ waren (C), einen gleichmäßigen Verlauf der Lesetiefe ohne auffälligen Abfall nach 349 Nukleotiden.

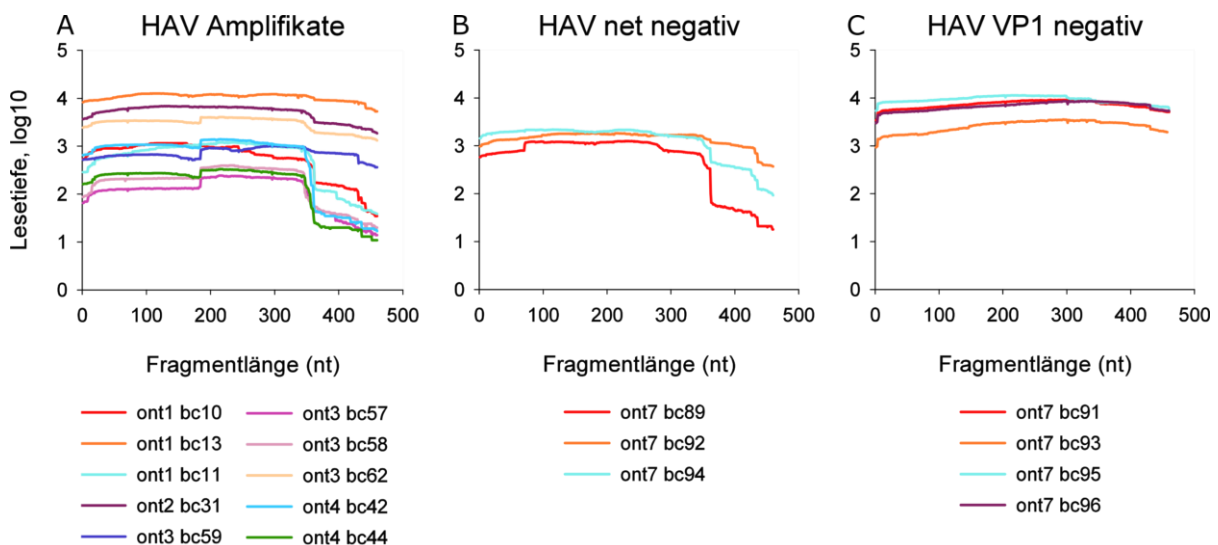


Abbildung 12: Verläufe der Lesetiefe über das HAV VP1- und net-Fragment.

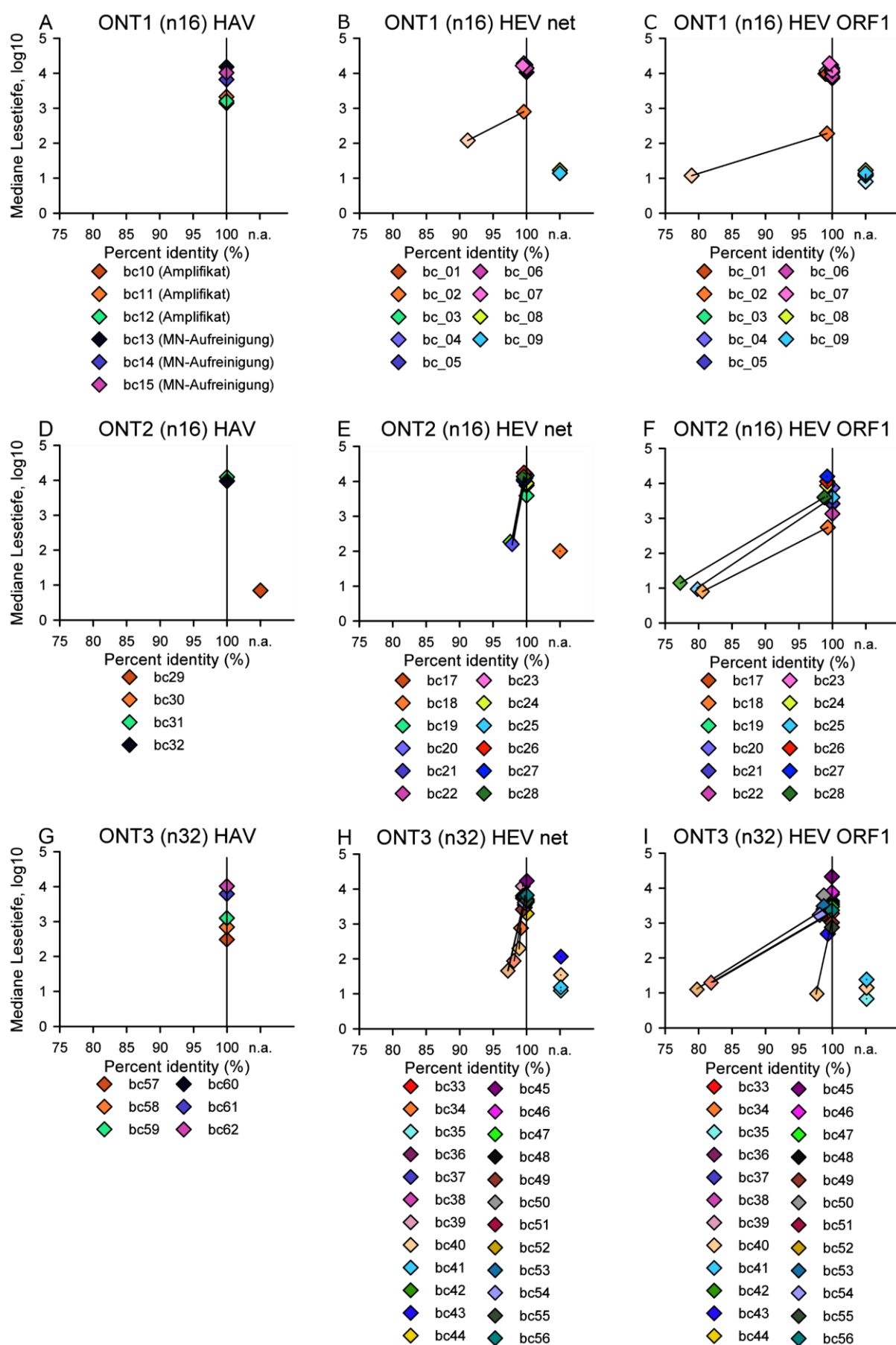
In (A) sind HAV-Amplifikate dargestellt, die sowohl VP1- als auch net-positiv sind. (B) zeigt VP1-positive/net-negative Proben und (C) ausschließlich net-positive, also VP1-negative Proben. Die x-Achse zeigt die Fragmentlänge in Nukleotiden an. Auf der y-Achse wird die logarithmierte Lesetiefe beschrieben. bc: Barcode; log: dekadischer Logarithmus; ont: Oxford Nanopore Technologies; nt: Nukleotide

3.6 Zusammenhang zwischen der Referenzsequenz, der medianen Lesetiefe und der Percent Identity

Ein Vergleich der ONT-Konsensussequenzen mit den entsprechenden Sanger-Sequenzen offenbarte für eine Subgruppe der Proben eine Percent Identity von nahezu 100 %. Demgegenüber wiesen einzelne, durch Sanger als positiv identifizierte Proben eine sehr niedrige mediane Lesetiefe (zum Teil unter 100 Reads) und eine deutlich reduzierte Percent Identity auf. Eine detaillierte Analyse ergab, dass ausschließlich Proben der Subtypen 3c (HEV) und IA (HAV) eine nahezu vollständige Übereinstimmung erzielten.

Infolge dieser Beobachtungen wurde der bioinformatische Workflow angepasst. Das Mapping erfolgte nun nicht mehr wie initial ausschließlich an die Referenzsequenz des dominanten Subtyps (3c für HEV, IA für HAV), sondern an ein repräsentatives Set verschiedener Referenzsequenzen. Diese Umstellung ermöglichte bei zuvor nicht auswertbaren Proben die erfolgreiche Erstellung von Konsensussequenzen mit hoher Übereinstimmung und deutlich verbesserten Lesetiefen. Diese Ergebnisse belegen, dass die Mapping-Rate (und daraus resultierend sowohl die Lesetiefe als auch die prozentuale Identität zur Sanger-Sequenz) stark von der gewählten Referenzsequenz abhängt.

Weiterhin zeigte sich ein systematischer Unterschied zwischen HAV- und HEV-positiven Proben: Die HAV-Proben zeigten fast durchweg eine Konzentration der Percent-Identity-Werte bei 100 %, während HEV-Proben mit Werten zwischen 98 % und 99,79 % breit streuten, was auf die höhere genetische Variabilität hindeutet (siehe Abbildung 13).



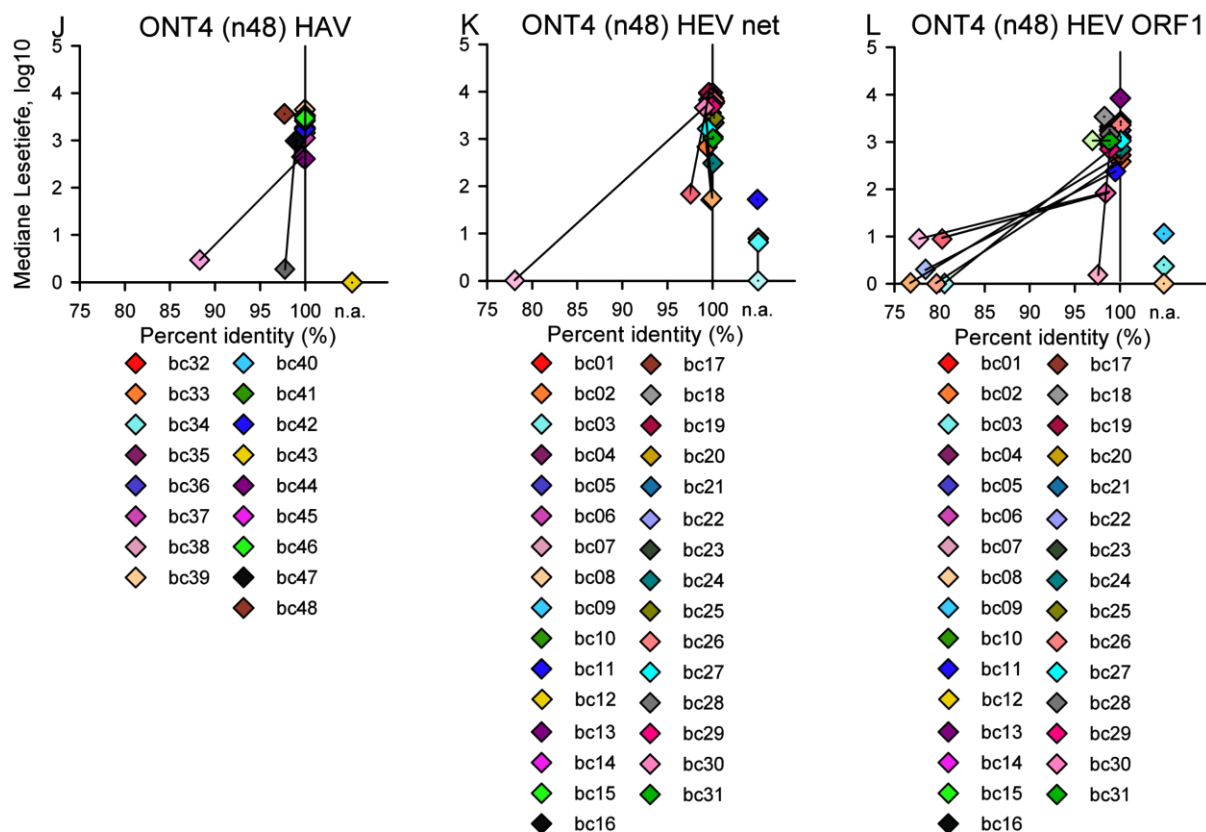


Abbildung 13: Zusammenhang zwischen der gemappten Referenzsequenz, Lesetiefen und Percent Identity.

Pro Sequenzierlauf und pro Fragment ist jeweils ein Diagramm dargestellt (A-L). Diese zeigen auf der x-Achse die Percent Identity der generierten Konsensussequenz zur Sanger-Sequenz und auf der y-Achse die zugehörige logarithmierte Lesetiefe der Probe. Negative Proben, bei denen keine Percent Identity berechnet werden konnte, sind unter n.a. aufgeführt. Die pastellfarbenen Diamanten repräsentieren die Ergebnisse der initialen Auswertung mit nur einer Referenzsequenz. Die Verbindungslinie zum Diamanten in kräftigerer Farbe stellt (falls vorhanden) die Verbesserung der Ergebnisse durch das Mapping an ein repräsentatives Set an Referenzsequenzen dar.

bc: Barcode; log: dekadischer Logarithmus; MN: Macherey-Nagel; ONT: Oxford Nanopore Technologies; n: Anzahl; n.a.: nicht anwendbar; nt: Nukleotide

3.7 Validierung negativer Proben sowie Betrachtung der Anzahl der Nukleotide und der durchschnittlichen Lesetiefe als Qualitätskriterien

Ergänzend zur Auswertung der Percent Identity und der Lesetiefen wurde eine phylogenetische Analyse durchgeführt, um die Herkunft der Sequenzfragmente aus den eigentlich negativen Proben zu überprüfen. Dabei zeigte sich, dass die detektierten, teilweise unvollständigen viralen Sequenzen aus Proben mit äußerst

geringen Lesetiefen vereinzelt eng mit den Sequenzen bestätigter positiver Proben desselben Laufs verwandt waren. Eine einzelne, dominante Kontaminationsquelle ließ sich dabei nicht identifizieren, was auf sporadische lauf-interne Verschleppungen (Cross-Contamination) hindeutet (siehe Anhang, 7b. Phylogenetische Analysen).

Abbildung 14 zeigt die Anzahl der N-Nukleotide auf der x-Achse (unbestimmte Basen) in den generierten Konsensussequenzen im Verhältnis zur medianen Lesetiefe in logarithmischer Skalierung (y-Achse). Obwohl negative Proben (als rote Dreiecke dargestellt) meist eine geringere Lesetiefe als die positiven Proben (als grüne Kreise dargestellt) aufwiesen, variierte deren N-Anzahl stark von 0 bis über 400. Somit wiesen einige negative Proben trotz minimaler Datenmenge keine N-Nukleotide auf, während gleichzeitig einige positive Proben (grüne Punkte) trotz hoher Lesetiefen eine zwar geringe, aber dennoch vorhandene Anzahl von N-Nukleotiden (von 0–8) aufwiesen. Diese Ergebnisse belegen, dass die Anzahl der N-Nukleotide kein verlässliches Kriterium zur Unterscheidung positiver und negativer Proben darstellt. Somit konnte die ursprüngliche Annahme, dass eine erhöhte N-Anzahl konsistent mit falsch-negativen Ergebnissen oder geringer Datenqualität korreliert, nicht bestätigt werden.

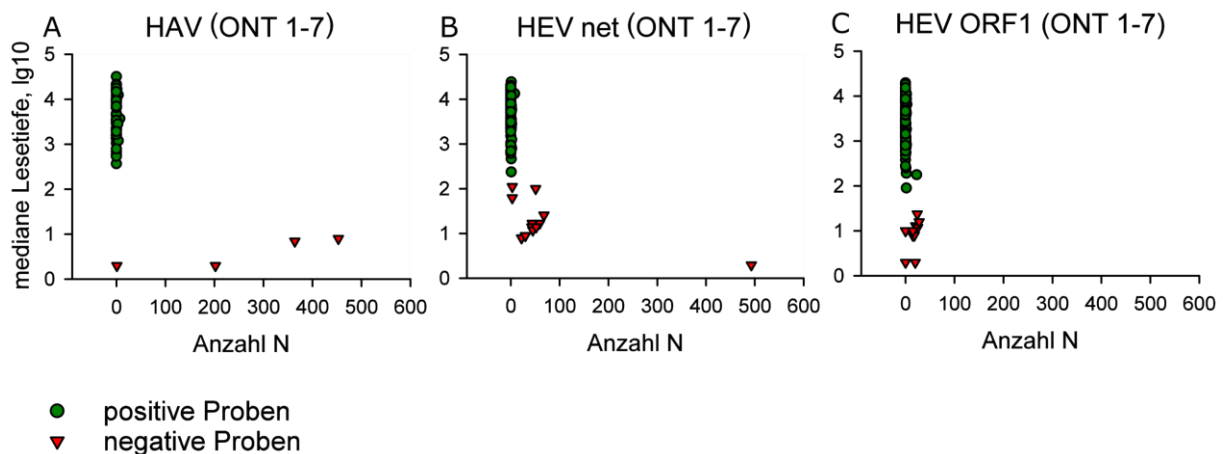


Abbildung 14: Zusammenhang zwischen der Anzahl der N-Nukleotide und der Lesetiefe.

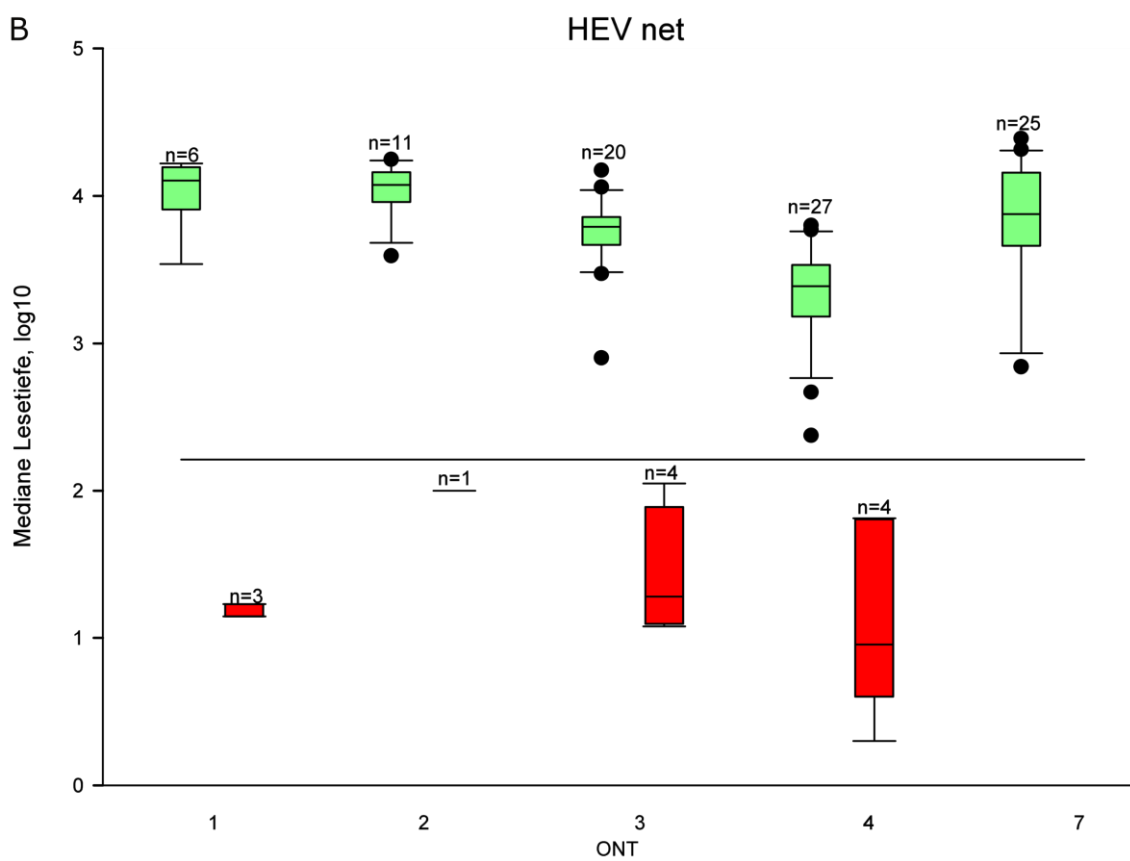
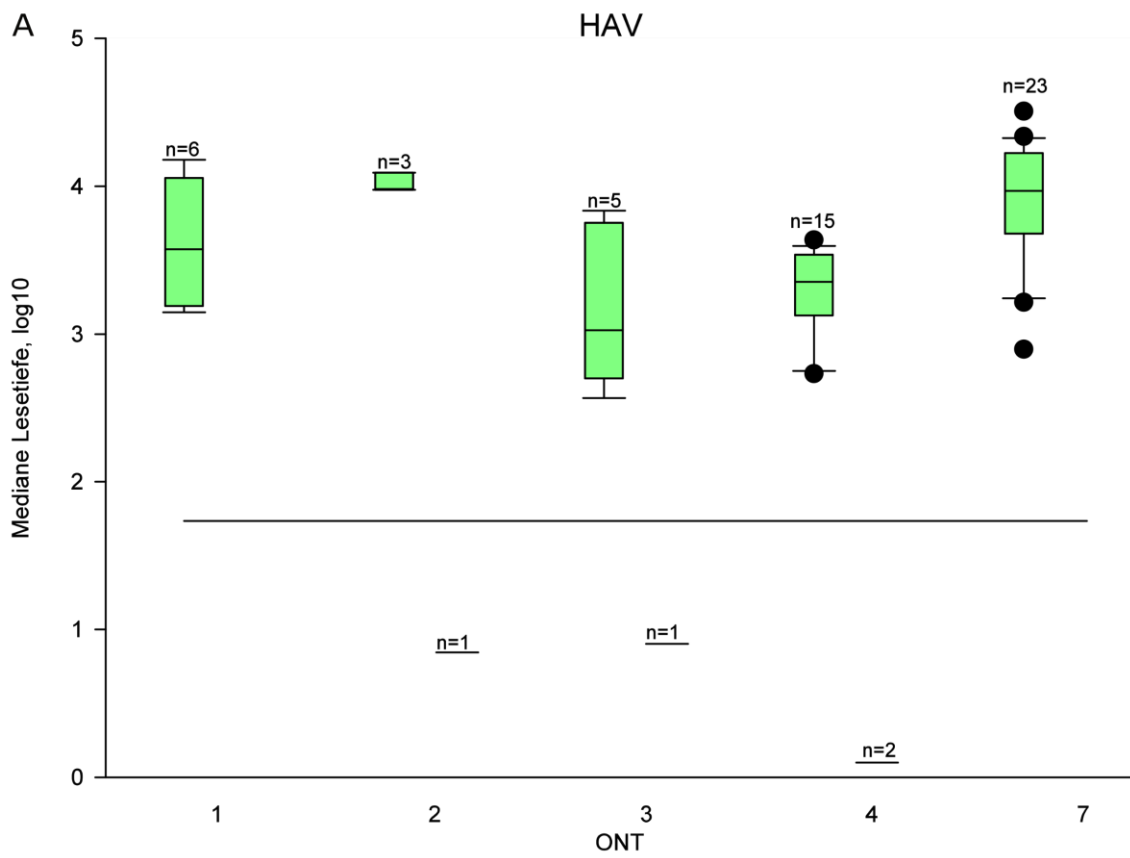
Auf der x-Achse ist die Anzahl der N-Nukleotide in der generierten Konsensussequenz abgebildet. Die y-Achse zeigt die zugehörige mediane Lesetiefe logarithmiert an. Positive Proben sind mit einem grünen Kreis eingezeichnet, während negative Proben mit einem roten Dreieck dargestellt sind.

log: dekadischer Logarithmus; N: N-Nukleotide; ONT: Oxford Nanopore Technologies

Im Gegensatz zur Anzahl der unbestimmten Nukleotide erwies sich die mediane Lesetiefe als robuster Qualitätsparameter. Die Analyse über alle Fragmente und

Sequenzierläufe hinweg zeigte, dass positive Proben konsistent höhere Lesetiefen erzielten als negative Proben, was die entsprechenden Boxplot-Analysen in Abbildung 15 verdeutlichen. Fragment-spezifisch betrachtet war in sämtlichen Läufen die maximale mediane Lesetiefe der negativen Proben stets geringer als die minimale mediane Lesetiefe der positiven Proben.

Um einen Schwellenwert zur Abgrenzung zu ermitteln, wurde zunächst ein Fragment-übergreifender logarithmierter Mittelwert aus den medianen Lesetiefen aller positiven und negativen Proben berechnet. Dieser lag bei 2,0115 (103 Reads). Es zeigte sich jedoch, dass dieser universelle Wert nicht ausreichte, um positive und negative Proben innerhalb einzelner Läufe über alle drei Zielregionen (HEV net, HEV ORF1 und HAV) hinweg fehlerfrei zu differenzieren. Stattdessen konnten Fragment-spezifische, Lauf-übergreifende logarithmierte Grenzwerte für eine konsistente Differenzierung definiert werden. Diese berechnen sich aus dem Mittelwert der höchsten Lesetiefe einer negativen Probe und der niedrigsten Lesetiefe einer positiven Probe innerhalb des jeweiligen Fragments: Für das HEV net-Fragment betrug dieser Cut-off-Wert 2,212 (163 Reads), für das HEV ORF1-Fragment 1,667 (46 Reads) und für HAV 1,7345 (54 Reads). Diese Fragment-abhängigen Differenzen legen nahe, dass ein pauschaler Mittelwert für alle Fragmente zu Fehlklassifizierungen führen würde und somit eine differenzierte Betrachtung der Amplikons für die Validierung der Ergebnisse zwingend erforderlich ist.



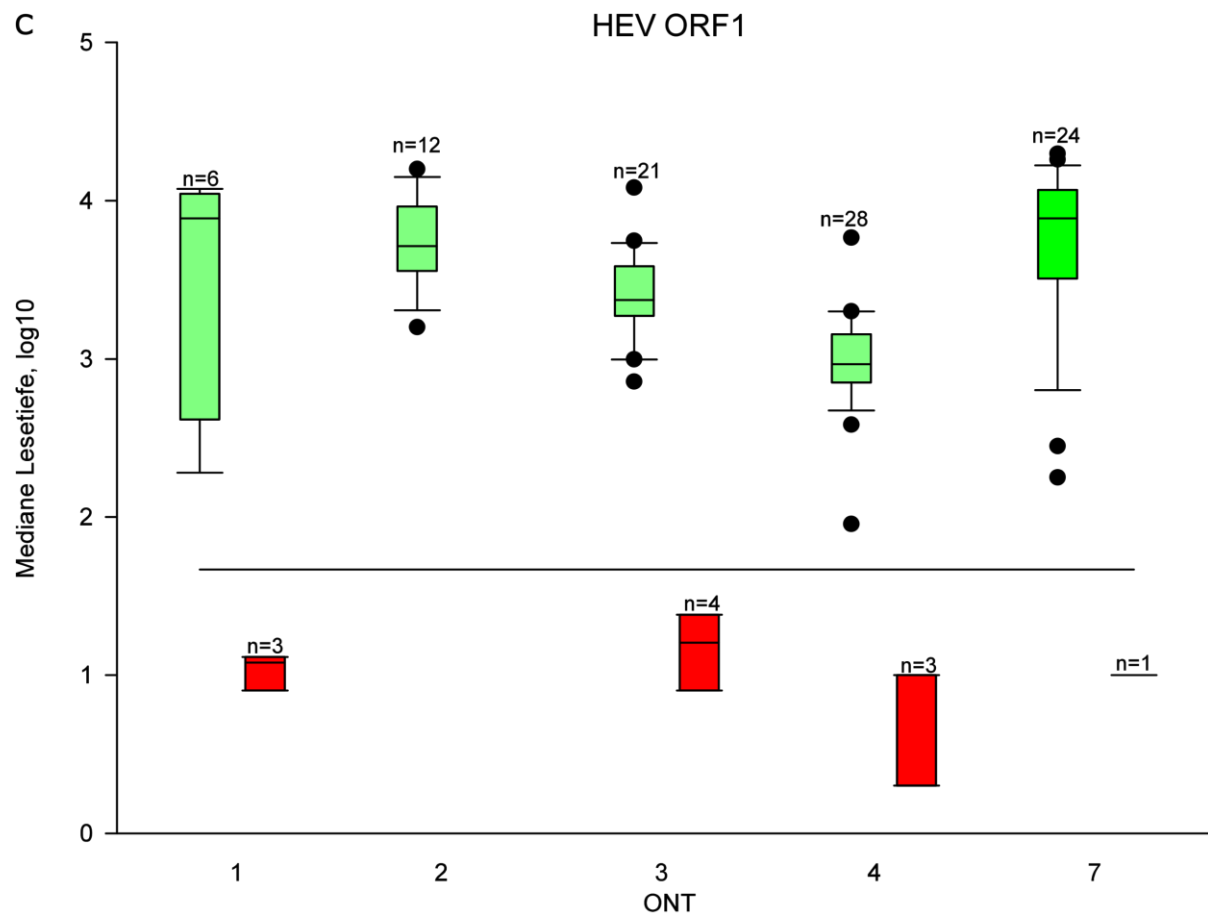


Abbildung 15: Verteilung der Lesetiefen positiver und negativer Proben.

Auf der x-Achse sind die einzelnen Sequenzierläufe dargestellt. Die y-Achse zeigt die mediane logarithmierte Lesetiefe. Die Lesetiefen der positiven Proben sind als grüne Boxplots dargestellt, während die negativen Proben in rot abgebildet sind. Die horizontale Linie markiert den laufübergreifenden, Fragment-spezifischen Cut-off-Wert, der sich aus dem Mittelwert der höchsten Lesetiefe der negativen Proben und der niedrigsten Lesetiefe der positiven Proben innerhalb des jeweiligen Fragments errechnet.

log: dekadischer Logarithmus; n: Probenanzahl; ONT: Oxford Nanopore Technologies

4 Diskussion

Das Ziel der vorliegenden Arbeit war die Entwicklung eines qualitativ gleichwertigen Sequenzierverfahrens für HAV- und HEV-Fragmente als Alternative zur bewährten Sanger-Sequenzierung. Ein zentraler Fokus lag dabei auf der Automatisierbarkeit, um den manuellen Laboraufwand signifikant zu reduzieren. Durch die Kombination einer amplikonbasierten Nanopore-Sequenzierung mit einer eigens entwickelten, vollständig automatisierten Bioinformatik-Pipeline wurde ein Workflow implementiert, der 262 Proben aus fünf Sequenzierläufen erfolgreich prozessierte. Für alle positiven Proben konnten dabei Konsensussequenzen mit einer Übereinstimmung von 98–100 % gegenüber der Sanger-Sequenzierung generiert werden. Der Workflow erwies sich damit sowohl hinsichtlich seiner Aussagekraft als auch seiner praktischen Umsetzbarkeit als geeignete Alternative zur Sanger-Sequenzierung.

Die Analysen zeigen, dass die mediane Lesetiefe ein verlässlicher Parameter für die Qualität der generierten Konsensussequenz ist. Wie erwartet wiesen positive Proben für alle untersuchten Fragmente höhere Lesetiefen auf als negative Proben. Ein Lauf- und Fragment-übergreifender Grenzwert erwies sich als ungeeignet, da dieser zu Fehlklassifikationen führen kann. Fragment-spezifische Cut-off-Werte (logarithmierte Werte von 2,212 für HEV net, 1,667 für HEV ORF1 und 1,7345 für HAV) ermöglichten hingegen eine eindeutige Differenzierung. Die beobachteten Unterschiede zwischen den Fragmenten könnten auf variierende Amplifikationseffizienzen oder unterschiedliche Sequenzkomplexitäten zurückzuführen sein.

Laut Riaz et al. (2021) (74) sind bei der Nanopore-Technologie zur Vollgenomsequenzierung des Hepatitis-C-Virus (HCV) für eine vergleichbare Konsensussequenz wie bei der Illumina-Sequenzierung über 300 Reads erforderlich. Wenngleich ein direkter Vergleich aufgrund der unterschiedlichen Methodik (Vollgenom- vs. Amplikon-Sequenzierung) nur bedingt möglich ist, verdeutlichen die Ergebnisse dennoch, dass für eine verlässliche Konsensussequenz zielgerichteter Amplikons bereits im Durchschnitt 46–163 Reads (je nach Fragment) ausreichend sind. Dies ist insbesondere im Hinblick auf den diagnostischen Anwendungszweck dieser Arbeit relevant. Die Lesetiefe stellt dabei einen zentralen Qualitätsparameter

zur Bewertung der Sequenzqualität dar, da sie Rückschlüsse auf die Genauigkeit der Basenaufrufe und die Verlässlichkeit der erstellten Konsensussequenz erlaubt. Eine ausreichende Lesetiefe verbessert die Qualität der Konsensusbildung, reduziert Sequenzierfehler und erhöht die Detektionsrate seltener Mutationen (75). Des Weiteren sinkt mit höherer Lesetiefe die Fehlerrate signifikant (76,77).

Auch die Studie von Batista et al. (2020) (78) untermauert die vorliegenden Ergebnisse. Wenngleich die Autoren einen PCR-freien Ansatz verwendeten und damit einen methodisch anderen Weg einschlugen als die vorliegende Amplikon-basierte Sequenzierung, zeigen ihre Ergebnisse ebenfalls, dass die ONT-Technologie zur zuverlässigen HAV-Charakterisierung auch auf Fragmentebene geeignet ist. So konnten vollständige HAV-Genome sowie Teilfragmente wie VP1 mit hoher Übereinstimmung zu Referenzgenomen reproduzierbar detektiert werden. Somit stimmen die Befunde dieser Studie mit den vorliegenden Ergebnissen überein und bestätigen die Übertragbarkeit der ONT-Technologie auf den amplikonbasierten diagnostischen Einsatz.

Noch näher am vorliegenden Versuchsaufbau ist die Arbeit von Lee et al. (2023) (79), die eine amplikonbasierte Nanopore-Sequenzierung zur molekularen Diagnostik von HAV einsetzt. Der Fokus dieser Studie lag jedoch auf der Erstellung hochqualitativer Konsensussequenzen selbst bei sehr niedrigen Viruskonzentrationen (10–100 Kopien/µl), wozu erheblich höhere Lesetiefen (6.000–50.000) nötig waren. Die vorliegenden Ergebnisse zeigen, dass für die gezielte Unterscheidung positiver und negativer Proben deutlich geringere Lesetiefen ausreichend sind. Die Studie von Lee et al. bestätigt damit vor allem die prinzipielle Leistungsfähigkeit der ONT-Technologie, während die Ergebnisse dieser Arbeit deren diagnostisches Potenzial unter ressourcenschonenden Bedingungen (technisch und praktisch) verdeutlichen.

Die Anzahl der N-Nukleotide erwies sich hingegen als unzuverlässiger Qualitätsparameter, da sowohl positive als auch negative Proben inkonsistente Werte aufwiesen. In Einzelfällen konnten aus negativ klassifizierten Proben nahezu vollständige Konsensussequenzen generiert werden.

Als mögliche Ursache kommen laborbedingte Artefakte oder Kontaminationen in Betracht. Wie Zinter et al. (2019) (80) für metagenomische Sequenzierungsexperimente systematisch beschreiben, können diese an jedem Punkt

des Workflows durch Umgebung, Reagenzien, Laborpersonal oder Geräte eingetragen werden. Besonders kritisch wird das bei Proben mit geringer Nukleinsäure-Inputmenge, bei denen Kontaminationen einen überproportional hohen Anteil der sequenzierten Reads ausmachen können. Dass solche Kontaminationen auch unter realen Laborbedingungen praktische Relevanz besitzen, veranschaulicht ein internationaler Ringversuch von Baylis et al. (2021) (81) mit molekularen Typisierungsmethoden für HEV. In dieser Studie wurde in mehreren Laboren eine Kontamination negativer Kontrollproben mit HEV-Sequenzen beobachtet, entweder durch Kreuzkontamination innerhalb des Panels oder durch Kontamination mit HEV-positivem Material, das bereits im Labor vorhanden war. Die Sequenzen wiesen eine vollständige Übereinstimmung mit bekannten Zellkulturstämmen auf, was auf eine Kontamination durch virales Ausgangsmaterial, Plasmide oder amplifizierte Nukleinsäuren hindeutet. Zwar kamen im Ringversuch von Baylis et al. (2021) unterschiedliche Sequenzierungsplattformen zum Einsatz, darunter Illumina, Sanger und ONT, jedoch zeigte sich plattformübergreifend, dass selbst geringe Kontaminationen ausreichen können, um in ursprünglich negativen Proben virale Sequenzen nachzuweisen. Diese Beobachtungen unterstreichen die Bedeutung standardisierter Arbeitsbedingungen und Kontaminationskontrollen unabhängig von der verwendeten Technologie, insbesondere in Laboren mit regelmäßigem Kontakt zu positivem Material, um die Bildung artifizieller Sequenzen zu vermeiden. Vor diesem Hintergrund erscheint es plausibel, dass auch in der vorliegenden Studie die niedrigen, aber dennoch vorhandenen Reads in ursprünglich negativen Proben auf laborinterne Hintergrundkontaminationen zurückzuführen sind.

Darüber hinaus ergab die Analyse der Verläufe der Lesetiefen über das HAV-Genom charakteristische Muster: Ein Einbruch nach dem 349. Nukleotid im net-Fragment war bei net- und VP1-positiven Proben sowie bei nur VP1-positiven Proben erkennbar, nicht jedoch bei nur net-positiven Proben. Dieses Muster legt nahe, dass durch die Überlappung der beiden Fragmente VP1 und net in diesem Abschnitt das Lesetiefen-Profil beeinflusst wird. Die Ausnahmeprobe des Barcodes 92, die keinen ausgeprägten Einbruch zeigte, könnte auf eine geringere Sensitivität der Sanger-Sequenzierung im Vergleich zur Nanopore-Sequenzierung hinweisen, die in diesem Fall auch schwache Signale detektieren konnte. Es könnte auch auf technische Besonderheiten wie eine

möglicherweise lineare statt einer exponentiellen Amplifikation hinweisen. Insgesamt verdeutlichen diese Muster, dass die Analyse der Lesetiefenverläufe wertvolle Hinweise auf technische Besonderheiten einzelner Proben liefern kann.

Die Wahl der Referenzsequenz beim Mapping zeigte sich als zentraler Faktor für die Qualität der Konsensusbildung. Ein unpassendes Referenzgenom führte zu einer deutlich reduzierten Mapping-Rate, die ihrerseits eine ungenaue Konsensussequenz zur Folge hatte, was sich in einer niedrigen prozentualen Übereinstimmung mit der Sanger-Sequenz äußerte. Durch die Erweiterung des Referenzsequenzpanels um verschiedene HAV- und HEV-Subtypen konnte die Mapping-Rate signifikant verbessert werden. Dieser Zusammenhang wurde auch in der Studie von Batista et al. (2020) (78) beobachtet, in der erst mit der Referenzsequenz des korrekten HAV-Stammes eine Übereinstimmung der Konsensussequenz von 99,8 % erreicht werden konnte. Liu et al. (2016) (82) demonstrierten am Beispiel des Hepatitis-B-Virus, dass die Nutzung einer referenzfernen Sequenz beim Mapping zu einer drastisch geringeren Abdeckung, einer geringeren Read-Zuordnung und einer Vielzahl falsch identifizierter Varianten führte. Die Autoren fordern daher explizit die Nutzung probenspezifischer oder zumindest genotypnaher Referenzsequenzen, um Analyseverzerrungen zu vermeiden. Dieser Zusammenhang ist direkt auf HEV übertragbar, das ebenfalls eine hohe genetische Variabilität zwischen den Subtypen aufweist. Ein weiteres anschauliches Beispiel liefert die Studie von Lapointe et al. (2021) (83) zur Sequenzierung des HCV-Genoms. Auch hier führte ein starres Mapping gegen eine nicht passende Referenz zu unvollständiger Abdeckung und fehlerhafter Genotypzuweisung. Erst ein iterativer Ansatz mit multiplen Referenzgenomen und dynamischer Anpassung ermöglichte eine vollständige und akkurate Assemblierung. Dabei wurde in jeder Iteration ein automatischer Genotyp-Check durchgeführt, um sicherzustellen, dass die gewählte Referenz zum dominanten Subtyp passte. Dieses Verfahren verhinderte Fehlzuordnungen und unterstreicht auch für HAV- und HEV-Analysen die Relevanz einer adaptiven Referenzstrategie zur Vermeidung von Mapping-Bias. Die Relevanz dieser Problematik wird schließlich auch durch die Ergebnisse von Zsichla et al. (2024) (84) zur HIV-1-Genomassemblierung bestätigt. Hier zeigte sich, dass bei Verwendung einer falschen Referenzsequenz selbst etablierte Analysepipelines deutlich schlechtere Ergebnisse, wie beispielsweise

geringere Konsensusqualität und abweichende Variantenerkennung, lieferten. Nur bei genetisch nah verwandten Referenzen wurden durchgängig robuste und konsistente Ergebnisse mit den getesteten Pipelines erzielt.

Ein möglicher nächster Schritt zur methodischen Weiterentwicklung könnte darin bestehen, künftig gegen eine möglichst umfassende Referenzdatenbank, etwa die vollständige GenBank-Datenbank mit allen bekannten HAV- und HEV-Subtypen, zu mappen. Auf diese Weise ließe sich automatisiert die jeweils passendste Referenzsequenz identifizieren und so die Konsensusqualität weiter optimieren. Allerdings ginge dieser Ansatz mit deutlich längeren Rechenzeiten während des Mappings einher, da der Verarbeitungsaufwand mit der Anzahl der Referenzsequenzen erheblich ansteigt. Deshalb erscheint es in der Praxis sinnvoller, nur eine begrenzte, aber repräsentative Auswahl an Referenzsequenzen zu verwenden, um den Kompromiss zwischen Genauigkeit und Rechenaufwand zu optimieren. Ein weiterer Ansatzpunkt zur Optimierung des bioinformatischen Workflows liegt im gezielten Vergleich unterschiedlicher Tools zur Generierung der Konsensussequenz. Für das Mapping wurde in der vorliegenden Arbeit *minimap2* (60) verwendet, das aktuell als einer der leistungsfähigsten Mapper gilt. Dennoch ist zu beachten, dass die anschließende Wahl des Softwaretools zur Konsensusbildung (z. B. *samtools*, *iVar*, *Medaka*) einen nicht zu unterschätzenden Einfluss auf die Qualität und Genauigkeit der endgültigen Konsensussequenz haben kann. Aufgrund begrenzter zeitlicher Ressourcen wurde dieser Aspekt im Rahmen der vorliegenden Arbeit nicht weiter vertieft. Er bietet jedoch großes Potenzial für zukünftige Analysen, um die Konsensusqualität und Auswertungssicherheit noch weiter zu steigern.

Ein Erklärungsansatz für die beobachteten Unterschiede in der Übereinstimmung der Nanopore-Konsensussequenzen zu den Sanger-Konsensussequenzen zwischen HAV- und HEV-Proben liegt in der genetischen Diversität. HAV-Proben weisen im Vergleich zu anderen RNA-Viren eine relativ geringe genetische Variabilität auf (85,86). Dies wird unter anderem durch das begrenzte Wirtsspektrum, welches hauptsächlich auf den Menschen beschränkt ist, sowie die fehlende zoonotische Übertragung begünstigt (87). Laut Cristina und Costa-Mattioli (2007) (86,88) ist HAV genomisch stabil mit einer signifikant niedrigeren Mutationsrate als viele andere RNA-Viren, was sich in einer vergleichsweise homogenen Populationsstruktur und geringen

Sequenzdivergenzen niederschlägt (86,88). Auch Smith und Simmonds (2018) (89) unterstreichen, dass HAV innerhalb der Gattung *Hepatovirus* lediglich eine moderate Diversität aufweist. Im Gegensatz dazu ist HEV deutlich variabler. Verschiedene Studien belegen, dass HEV durch ein breites Wirtsspektrum, häufige zoonotische Übertragungen sowie eine geografisch stark differenzierte Genotypenverteilung geprägt ist. Primadharsini et al. (2019) (90) beschreiben HEV als hochdiverses Virus, das sich kontinuierlich weiterentwickelt und an verschiedene tierische Wirte wie Schweine, Kaninchen und Wildtiere anpasst. Smith et al. (2013) (91) betonen ebenfalls die hohe genetische Diversität von HEV und leiten daraus die Notwendigkeit einer differenzierten molekularen Klassifikation ab. Darüber hinaus weist HEV, wie die meisten RNA-Viren, eine hohe Mutationsrate auf, was wesentlich auf die Abwesenheit eines Proofreading-Mechanismus der viralen RNA-Polymerase zurückzuführen ist. In Kombination mit einer hohen Replikationsrate (92) führt dies zu einer gesteigerten Fehlerrate, wodurch Mutationen entstehen, die sich über Generationen hinweg in der Viruspopulation akkumulieren können (33). Außerdem befindet sich auf dem ORF1-Fragment eine hypervariable Region (33,93), welche eine hohe Sequenzvariabilität besitzt, die vermutlich zur Anpassung an verschiedene Wirte beiträgt (93). Dieses Erkenntnis unterstreicht erneut die essentielle Bedeutung eines repräsentativen Referenzsequenzpanels für eine präzise Konsensusbildung.

Erwartungsgemäß zeigte sich zwischen der Probenzahl pro Lauf und der erzielten Sequenziertiefe ein inverses Verhältnis: Je mehr Proben, desto geringer die Lesetiefe pro Probe. Dies ist auf die begrenzte Anzahl aktiver Nanoporen auf der Flowcell zurückzuführen, deren Kapazität bei einer höheren Probenanzahl auf mehr Proben aufgeteilt werden muss. Obwohl eine größere Probenanzahl pro Lauf ökonomisch vorteilhaft ist, kann sie die Datenqualität einzelner Proben, insbesondere bei niedriger Viruslast oder suboptimaler Ausgangsqualität des Probenmaterials, erheblich beeinträchtigen. Auch Baym et al. (2015) (94) belegen in ihrem Protokoll zur kostengünstigen Multiplex-Sequenzierung kleiner Genome, dass die Anzahl der Proben in engem Zusammenhang mit der resultierenden Lesetiefe und der finalen Datenqualität steht. Wenngleich sich diese Ergebnisse auf bakterielle Genome beziehen, ist das zugrundeliegende Prinzip auf die virale Sequenzierung übertragbar, da auch hier die verfügbare Sequenzierkapazität der limitierende Faktor ist.

Von ONT liegen für Viren derzeit keine spezifischen Empfehlungen bezüglich der maximalen Probenanzahl oder einer minimalen medianen Lesetiefe vor. Für bakterielle Isolate wird jedoch bei Verwendung des identischen Kits (*Rapid Barcoding Kit*) empfohlen, nicht mehr als 24 Proben pro Lauf zu laden und mindestens eine 50-fache Abdeckung anzustreben. Bei Pilz- und Hefepräparaten wird die Probenanzahl aufgrund der größeren und komplexeren Genome sogar auf maximal acht Proben begrenzt, um eine ausreichende Abdeckung gewährleisten zu können (95). Diese Richtwerte unterstreichen die Notwendigkeit einer zielgerichteten Laufplanung, die individuell auf die Genomgröße und die erwartete Viruslast der Proben abgestimmt sein muss.

Zudem war der unterschiedliche Rückgang der Lesetiefen bei einer Erhöhung der Probenanzahl bei verschiedenen Fragmenten auffällig. Während die HEV-Fragmente vergleichsweise stabile Lesetiefen aufwiesen, verzeichnete insbesondere das HAV-Fragment einen deutlich stärkeren Einbruch. Diese Diskrepanz deutet darauf hin, dass neben der reinen Probenanzahl (Multiplexing-Effekt) Fragment-spezifische Faktoren die Sequenzierleistung maßgeblich beeinflussen. Wie Santos et al. (2025) (96) beschreiben, hängt die Effizienz der Nanopore-Sequenzierung stark vom jeweiligen genomischen Kontext ab, was die vorliegenden Beobachtungen stützt.

Ein möglicher Erklärungsansatz liegt in der unterschiedlichen Basenzusammensetzung der Zielsequenzen. Das eher GC-arme HAV-Fragment weist mit einem GC-Gehalt von ca. 39 % eine vergleichsweise AT-reiche Sequenz auf, während die HEV-Fragmente mit ca. 55 % deutlich GC-reicher sind. Solche Unterschiede können dazu führen, dass die Zielsequenzen während der Bibliotheksvorbereitung oder beim Durchgang durch die Nanopore benachteiligt werden. Obwohl der GC-Gehalt des HAV-Fragments laut Ross et al. (2013) (97) für sich genommen als stabil und gut sequenzierbar gilt, scheint im vorliegenden Fall unter kompetitiven Bedingungen, also bei steigender Multiplexität, ein relativer Bias zugunsten der GC-reicheren HEV-Fragmente aufzutreten. Dies legt nahe, dass unter Konkurrenzbedingungen selbst moderate Unterschiede in der Basenzusammensetzung ausreichen, um eine stochastische Benachteiligung AT-reicher Fragmente zu bewirken. Während beide Fragmente einzeln betrachtet gut zu sequenzieren sind, führt die Erhöhung der Plexität offenbar dazu, dass das chemisch

stabilere HEV-Fragment bevorzugt prozessiert wird.

Um einen drastischen Abfall der Lesetiefe bei steigender Probenanzahl zu kompensieren, könnten gezielte Anreicherungsstrategien eine praktikable Lösung bieten. Santos et al. (2025) (98) betonen in diesem Zusammenhang, dass eine ausreichende Sequenziertiefe eine wesentliche Voraussetzung für die Qualität der generierten Sequenz ist. So zeigen Gilpatrick et al. (2020) (96) mit der *Nanopore-Cas9-targeted sequencing*-Methode einen vielversprechenden Ansatz, der durch den Einsatz von Cas9-basierten Adaptern eine extrem hohe mediane Abdeckung von bis zu 675-fach ermöglicht. Dadurch ist eine hochspezifische Ansteuerung einzelner Loci bei gleichzeitig geringem DNA-Input möglich. Eine solche alternative zielgerichtete Anreicherung könnte insbesondere für AT-reiche Fragmente wie HAV unter Multiplex-Bedingungen eine Verbesserung der Sequenzierleistung bewirken.

Auch die DNA-Konzentration beeinflusste die Lesetiefe: Unverdünnte Proben lieferten tendenziell bessere Ergebnisse als standardisierte Verdünnungen. Diese Beobachtung deutet darauf hin, dass bei ausreichender Probenqualität eine Vorverdünnung verzichtbar ist, was nicht nur Zeit und Material spart, sondern auch potenzielle Probenverluste oder Kontaminationen durch zusätzliche Pipettierschritte minimiert.

Ein Vergleich gereinigter und unaufgereinigter Amplifikate verdeutlichte den Vorteil einer DNA-Aufreinigung: Drei von vier Proben zeigten eine signifikant höhere mediane Lesetiefe nach der Aufreinigung mittels Macherey-Nagel-Kits. Dies untermauert die Annahme, dass Verunreinigungen wie Salze, Proteine, Primer oder PCR-Rückstände, die typischerweise in unaufgereinigten Amplifikaten enthalten sind, die Funktion der Nanoporen stören und so die Sequenzierleistung verringern können. Solche Verunreinigungen können Poren physisch blockieren (*Pore clogging*) oder die Effizienz der Adapter-Ligation behindern, was sich in einer geringeren Read-Anzahl und reduzierter Abdeckung äußert (99). Weitere mögliche Ursachen für die beobachteten Diskrepanzen könnten in der heterogenen Fragmentqualität und suboptimalen Probenzusammensetzung liegen. In Übereinstimmung mit diesen Befunden empfiehlt ONT explizit, auf eine hohe DNA-Reinheit zu achten. Als optimaler Reinheitswert gilt laut NanoDrop-Messung ein Verhältnis 260/280 von etwa 1,8. Werte darunter deuten auf Proteinverunreinigungen, Werte über 2,2 auf mögliche organische Rückstände wie

Phenole oder Salze hin (100). Der Einsatz eines standardisierten Reinigungsverfahrens wie des Macherey-Nagel-Kits führt zu höherer Ausbeute, gleichmäßigeren Fragmenten und insgesamt stabileren Sequenzierbedingungen, da Nebenprodukte wie Primer effektiv entfernt werden. Vor diesem Hintergrund ist die Integration eines DNA-Reinigungsschritts vor der Bibliothekspräparation ausdrücklich zu empfehlen und sollte als Standardbestandteil des Protokolls etabliert werden. Ob dieser Effekt in gleicher Weise für HEV-Proben gilt, wäre in weiterführenden Untersuchungen zu prüfen.

Zusammenfassend steht mit dem entwickelten Workflow ein bioinformatisch vollautomatisiertes, modular aufgebautes System zur Verfügung, das demonstriert, dass eine amplikonbasierte Nanopore-Sequenzierung in Kombination mit einem automatisierten Auswertungssystem eine schnelle und zuverlässige Alternative zur Sanger-Sequenzierung darstellt. Die finale Auswahl der Subtyp-spezifischen Konsensussequenzen auf Basis der höchsten Lesetiefen erreichte bei sämtlichen HAV- und HEV-positiven Proben eine Übereinstimmung zwischen 98 – 100 % mit den Sanger-Sequenzen. Die vollständige Automatisierung der Datenauswertung reduzierte den Bearbeitungsaufwand erheblich und ermöglichte einen hohen Probendurchsatz. Somit belegen die Ergebnisse die diagnostische Verlässlichkeit des Workflows und eröffnen praxisrelevante Potenziale für die Anwendung in der molekularen Virusdiagnostik, sowohl im Laboralltag als auch im Rahmen der epidemiologischen Überwachung.

5 Zusammenfassung

Das Hepatitis-A-Virus (HAV) und das Hepatitis-E-Virus (HEV) sind weltweit die häufigsten Erreger akuter Virushepatitiden. Steigende Fallzahlen erfordern eine effiziente molekulare Typisierung. Die Sanger-Sequenzierung gilt zwar weiterhin als Goldstandard, stößt jedoch aufgrund des hohen manuellen Laboraufwands und eingeschränkter Automatisierungsmöglichkeiten zunehmend an ihre Grenzen, insbesondere bei steigendem Probendurchsatz. Ziel der Arbeit war die Entwicklung und Validierung eines bioinformatisch automatisierten Sequenzierworkflows mittels Oxford Nanopore Technologies als ressourceneffiziente Alternative für die klinische Routine.

Aus 99 klinischen Proben wurden 262 Sequenzen mit dem *Rapid Barcoding Kit 96* auf einem *MinION Mk1C* generiert. Die bioinformatische Auswertung erfolgte durch eine modular aufgebaute Pipeline (Bash, R und Python), die ein adaptives Mapping gegen ein repräsentatives Referenzset der Zielregionen (HAV net, HAV VP1, HEV net und HEV ORF1) durchführt. Zusätzlich wurden der Einfluss der Probenaufreinigung (*Macherey-Nagel-Kit*) sowie variierender DNA-Konzentrationen auf die Datenqualität untersucht.

Die generierten Konsensussequenzen zeigten eine hohe Übereinstimmung (> 98 %) mit den Sanger-Referenzdaten. Als zentraler Qualitätsparameter zur Differenzierung valider viraler Signale von Kontaminationen erwies sich die mediane Lesetiefe. Fragment-spezifische Cut-off-Werte wurden bei 54 Reads (HAV), 163 Reads (HEV net) und 46 Reads (HEV ORF1) definiert. Sowohl die Aufreinigung als auch der Einsatz höherer DNA-Mengen steigerten die Zuverlässigkeit der Ergebnisse signifikant, während die Anzahl unbestimmter Nukleotide (N-Calls) als unzuverlässiges Kriterium eingestuft wurde.

Der Workflow bietet eine zuverlässige, zeitsparende und hochdurchsatzfähige Alternative zur Sanger-Sequenzierung. Die Arbeit liefert damit einen technischen Leitfaden für die Implementierung der Nanopore-Sequenzierung sowie spezifische Qualitätskriterien, die maßgeblich zur Standardisierung und Qualitätssicherung in der molekularen Hepatitis-Diagnostik beitragen.

6 Summary

Hepatitis A virus (HAV) and Hepatitis E virus (HEV) are leading causes of acute viral hepatitis worldwide. Increasing incidence, particularly in resource-limited settings, highlights the need for molecular typing. While Sanger sequencing remains the gold standard, it is increasingly reaching its limits due to high manual workload and limited automation. This study aimed to develop and validate a bioinformatically automated sequencing workflow using Oxford Nanopore Technologies as a resource-efficient alternative for routine clinical diagnostics.

From 99 clinical samples, 262 sequences were generated using the Rapid Barcoding Kit 96 on a MinION Mk1C. Bioinformatic analysis was performed via a modular pipeline (Bash, R and Python) utilizing adaptive mapping against a representative reference set of target regions (HAV net, HAV VP1, HEV net, HEV ORF1). In addition, the impact of nucleic acid purification (Macherey-Nagel Kit) and varying DNA concentrations on data quality was evaluated.

The generated consensus sequences showed high concordance (> 98 %) with Sanger reference data. Median read depth was identified as a key quality parameter for distinguishing valid viral signals from contamination. Fragment-specific cutoff values were defined at 54 reads (HAV), 163 reads (HEV net) and 46 reads (HEV ORF1). Both purification and higher template concentrations significantly improved result reliability, while the number of undetermined nucleotides (N-calls) proved unreliable as a standalone quality criterion.

The workflow offers a reliable, time-saving and high-throughput alternative to Sanger sequencing. This study provides a methodological framework and specific quality criteria for implementing Nanopore sequencing, significantly contributing to the standardization and quality assurance of molecular hepatitis diagnostics.

7 Anhang

a. Abbildungsverzeichnis

Abbildung 1	- Schematische Darstellung des HAV-Genoms.	6
Abbildung 2	- Phylogenetische Darstellung der humanpathogenen Genotypen I – III.	7
Abbildung 3	- Schematische Darstellung des HEV-Genoms.	9
Abbildung 4	- Schematische Darstellung der Genomorganisation von HEV und HAV.	17
Abbildung 5	- Schematischer Ablauf der bioinformatischen Auswertung.	21
Abbildung 6	- Plot von Barcode 66 (ONT 7).	24
Abbildung 7	- Tabelle mit ausgewerteten Referenzsequenzen von Barcode 66 (ONT7) (Statistikdatei).	24
Abbildung 8	- Übersicht Workflow bioinformatische Auswertung von den Roh-POD5-Dateien bis zur Konsensussequenz.	26
Abbildung 9	- Mediane Lesetiefen in Abhängigkeit von der Probenanzahl.	30
Abbildung 10	- Vergleich der medianen Lesetiefen zwischen unaufgereinigten Amplifikaten und Macherey-Nagel-Aufreinigungen.	31
Abbildung 11	- Vergleich zwischen unverdünnten und verdünnten Proben.	32
Abbildung 12	- Verläufe der Lesetiefe über das HAV VP1- und net-Fragment.	33
Abbildung 13	- Zusammenhang zwischen der gemappten Referenzsequenz, Lesetiefen und Percent Identity.	36
Abbildung 14	- Zusammenhang zwischen der Anzahl der N-Nukleotide und der Lesetiefe.	37
Abbildung 15	- Verteilung der Lesetiefen positiver und negativer Proben.	40

b. Phylogenetische Analysen

In diesem Abschnitt werden die phylogenetische Einordnung aller generierten Sequenzen der drei untersuchten Fragmente (Abbildung A1–A3) dargestellt. Die Analyse umfasst den Abgleich mit repräsentativen Referenzsequenzen sowie den Ergebnissen der Sanger-Sequenzierung. Die generierten Sequenzen sind benannt nach Sequenzierlauf, Barcode, Virustyp und institutsinterner Probennummer, wobei die negativen Proben zur besseren Identifikation rot markiert sind. Die Sanger-Sequenzen tragen den Namenszusatz „Sanger“, die Beschriftung der Referenzsequenzen erfolgt im Format Virus-Subtyp, Isolat und Accession.

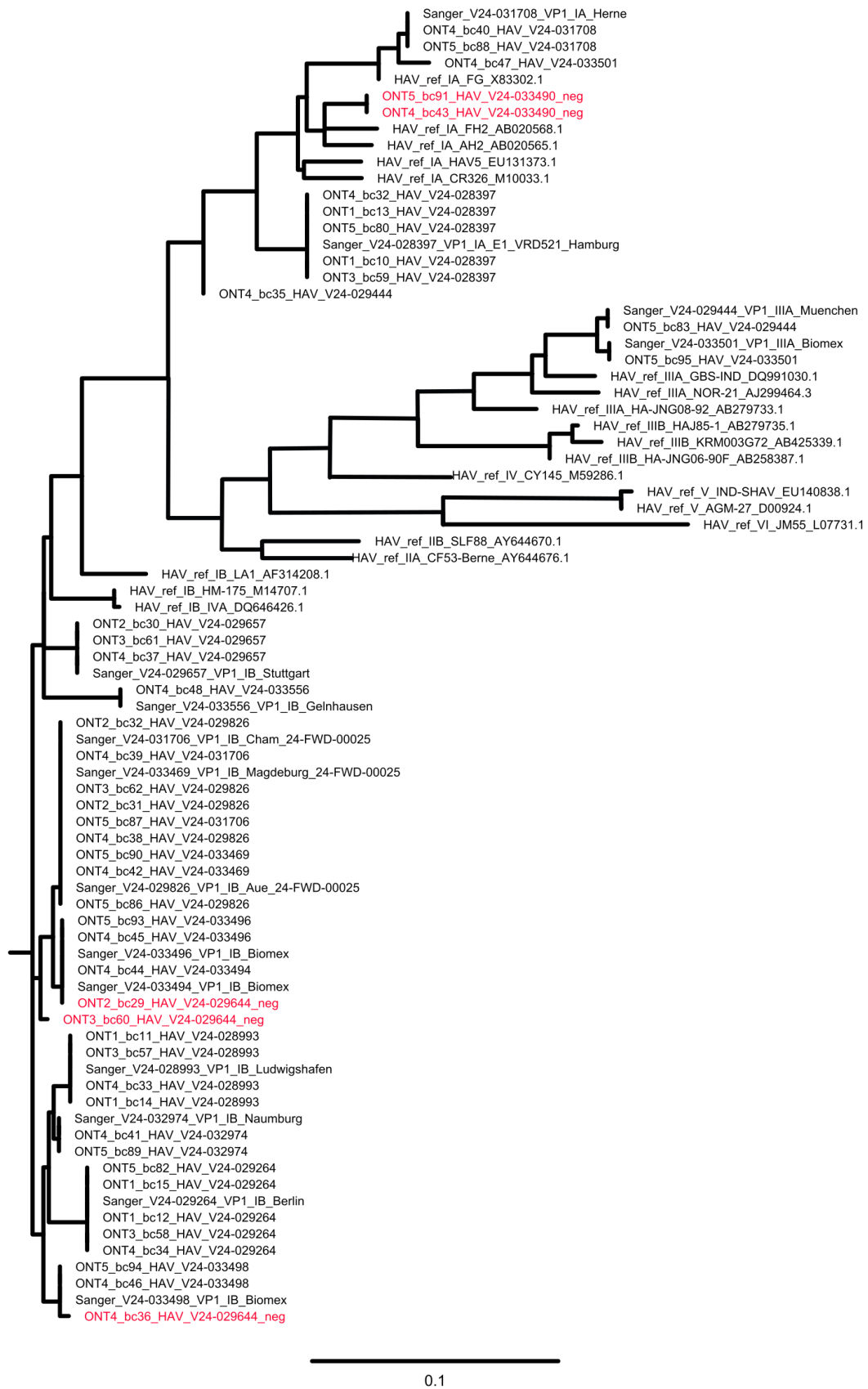


Abbildung A1: Phylogenetischer Baum von HAV.

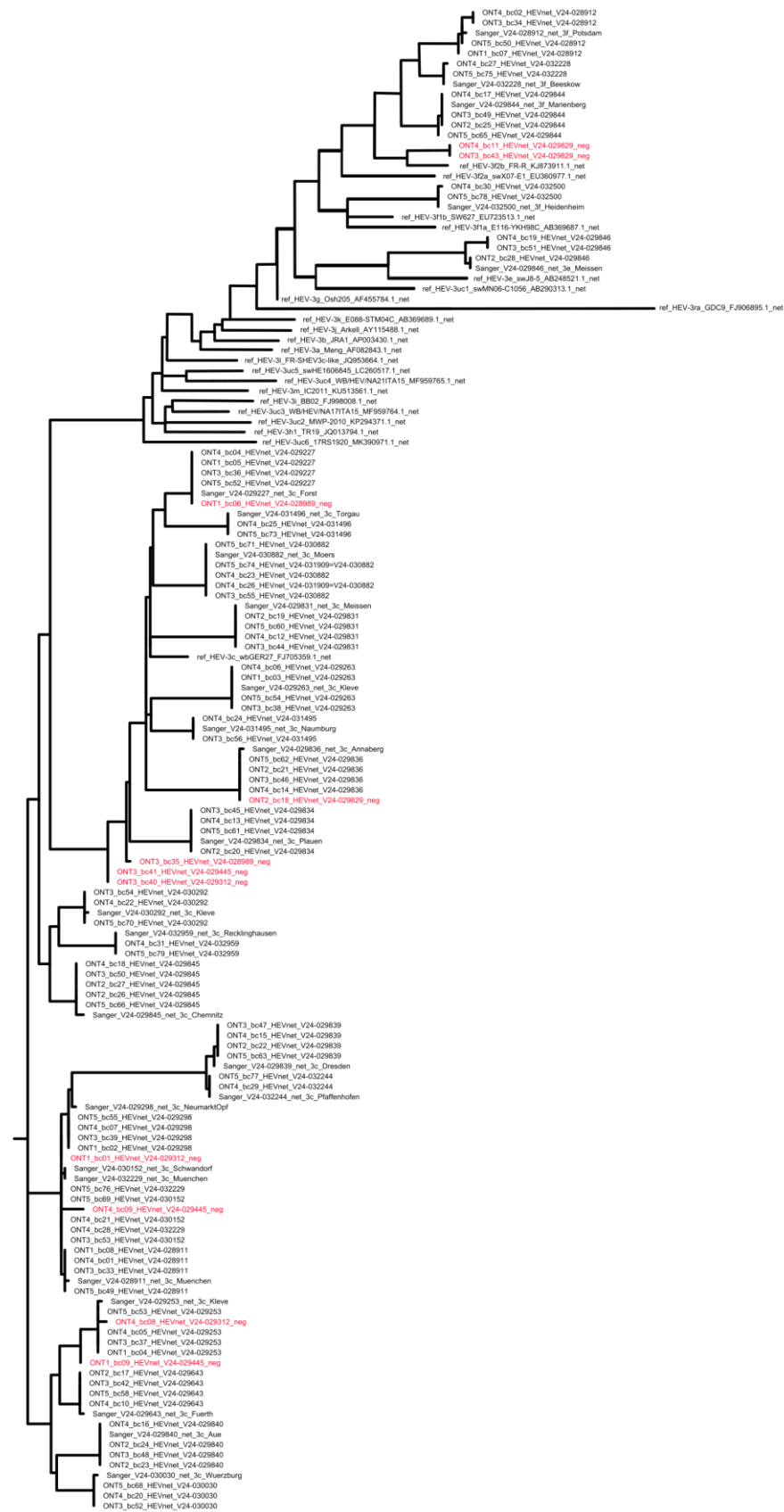


Abbildung A2: Phylogenetischer Baum von HEV net.

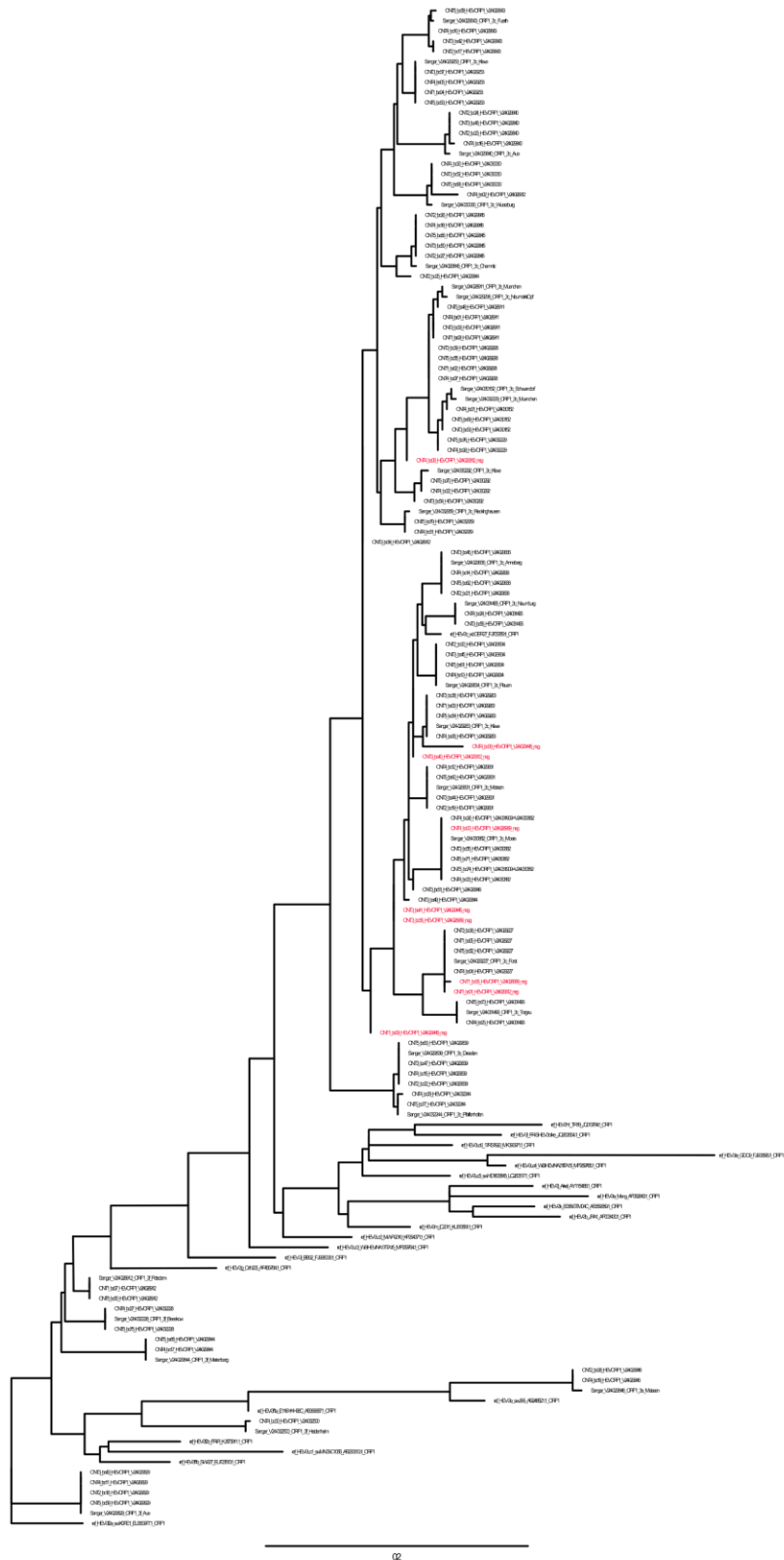


Abbildung A3: Phylogenetischer Baum von HEV ORF1.

c. Abkürzungsverzeichnis

A	Adenin
bp	Basenpaar
BR	Broad Range
BSA	Bovine Serum Albumin
C	Cytosin
CSV	Comma-separated values
DNA	Desoxyribonukleinsäure
et al.	Et alii, et aliae, et alia
G	Guanin
HAV	Hepatitis-A-Virus
HCV	Hepatitis-C-Virus
HEV	Hepatitis-E-Virus
HIV	Humanes Immundefizienz-Virus
HS	High Sensitivity
IMHR	Institut für Klinische Mikrobiologie und Hygiene am Universitätsklinikum Regensburg
kb	Kilobasen
log	Logarithmus
ml	Milliliter
MN	Macherey-Nagel
µl	Mikroliter
n	Anzahl
n. a.	Nicht anwendbar
ng	Nanogramm
NGS	Next-Generation Sequencing
nm	Nanometer
ONT	Oxford Nanopore Technologies
ORF	Open Reading Frame
PacBio	Pacific Biosciences
PCR	Polymerase Chain Reaction
PDF	Portable Document Format
pH	Pondus Hydrogenii
POD5	Portable Oxford Data version 5
RKI	Robert Koch-Institut
RNA	Ribonukleinsäure
RT	Raumtemperatur
SMRT	Single-Molecule Real-Time
T	Thymin
TSV	Tab-Separated Values
uc	Unclassified
USA	United States of America
UTR	Untranslated Region
VP	Viral Protein
WHO	Weltgesundheitsorganisation

8 Literaturverzeichnis

1. Miguères M, Lhomme S, Izopet J. Hepatitis A: Epidemiology, High-Risk Groups, Prevention and Research on Antiviral Treatment. *Viruses*. 2021;13(10). doi:10.3390/v13101900 Cited in: PubMed; PMID 34696330.
2. Cao G, Jing W, Liu J, Liu M. The global trends and regional differences in incidence and mortality of hepatitis A from 1990 to 2019 and implications for its prevention. *Hepatology International*. 2021;15(5):1068–82. doi:10.1007/s12072-021-10232-4 Cited in: PubMed; PMID 34345993.
3. Lemon SM, Ott JJ, van Damme P, Shouval D. Type A viral hepatitis: A summary and update on the molecular virology, epidemiology, pathogenesis and prevention. *J Hepatol*. 2017;68(1):167–84. doi:10.1016/j.jhep.2017.08.034 Cited in: PubMed; PMID 28887164.
4. Feinstone SM, Kapikian AZ, Purceli RH. Hepatitis A: detection by immune electron microscopy of a viruslike antigen associated with acute illness. *Science*. 1973;182(4116):1026–8. doi:10.1126/science.182.4116.1026 Cited in: PubMed; PMID 4356028.
5. Koff RS. Hepatitis A. *The Lancet*. 1998;351(9116):1643–9. doi:10.1016/S0140-6736(98)01304-X Cited in: PubMed; PMID 9620732.
6. Franco E, Meleleo C, Serino L, Sorbara D, Zaratti L. Hepatitis A: Epidemiology and prevention in developing countries. *World J Hepatol*. 2012;4(3):68–73. doi:10.4254/wjh.v4.i3.68 Cited in: PubMed; PMID 22489258.
7. Mohd Hanafiah K, Jacobsen KH, Wiersma ST. Challenges to mapping the health risk of hepatitis A virus infection. *International Journal of Health Geographics*. 2011;10(1):57. doi:10.1186/1476-072X-10-57 Cited in: PubMed; PMID 22008459.
8. Andani A, Mellou K, Dewda P, Eeuwijk J, Kassianos G, van Damme P, Steffen R. Evolution and Impact of Hepatitis A Epidemiology in Europe-Systematic Literature Review of the Last 20 Years. *J Viral Hepat*. 2025;32(1):e14030. doi:10.1111/jvh.14030 Cited in: PubMed; PMID 39526433.
9. Hofmeister MG, Xing J, Foster MA, Augustine RJ, Burkholder C, Collins J, McBee S, Thomasson ED, Thoroughman D, Weng MK, Spradling PR. Factors Associated With Hepatitis A Mortality During Person-to-Person Outbreaks: A Matched Case-Control Study-United States, 2016-2019. *Hepatology*. 2021;74(1):28–40. doi:10.1002/hep.31645 Cited in: PubMed; PMID 33217769.
10. Meyerhoff AS, Jacobs RJ. Transmission of hepatitis A through household contact. *J Viral Hepat*. 2001;8(6):454–8. doi:10.1046/j.1365-2893.2001.00317.x Cited in: PubMed; PMID 11703577.
11. Ndumbi P, Freidl GS, Williams CJ, Mårdh O, Varela C, Avellón A, Friesema I, Vennema H, Beebejaun K, Ngui SL, Edelstein M, Smith-Palmer A, Murphy N, Dean J, Faber M, Wenzel J, Kontio M, Müller L, Midgley SE, Sundqvist L,

- Ederth JL, Roque-Afonso A-M, Couturier E, Klamer S, Rebolledo J, Suin V, Aberle SW, Schmid D, Sousa R de, Augusto GF, Alfonsi V, Del Manso M, Ciccaglione AR, Mellou K, Hadjichristodoulou C, Donachie A, Borg M-L, Sočan M, Poljak M, Severi E. Hepatitis A outbreak disproportionately affecting men who have sex with men (MSM) in the European Union and European Economic Area, June 2016 to May 2017. *Euro Surveill.* 2018;23(33). doi:10.2807/1560-7917.ES.2018.23.33.1700641 Cited in: PubMed; PMID 30131095.
12. Mannucci PM, Gdovin S, Gringeri A, Colombo M, Mele A, Schinaia N, Ciavarella N, Emerson SU, Purcell RH. Transmission of hepatitis A to patients with hemophilia by factor VIII concentrates treated with organic solvent and detergent to inactivate viruses. The Italian Collaborative Group. *Ann Intern Med.* 1994;120(1):1–7. doi:10.7326/0003-4819-120-1-199401010-00001 Cited in: PubMed; PMID 7504424.
 13. Feng Z, Lemon SM. Peek-a-boo: membrane hijacking and the pathogenesis of viral hepatitis. *Trends Microbiol.* 2014;22(2):59–64. doi:10.1016/j.tim.2013.10.005 Cited in: PubMed; PMID 24268716.
 14. Gholizadeh O, Akbarzadeh S, Ghazanfari Hashemi M, Gholami M, Amini P, Yekanipour Z, Tabatabaie R, Yasamineh S, Hosseini P, Poortahmasebi V. Hepatitis A: Viral Structure, Classification, Life Cycle, Clinical Symptoms, Diagnosis Error, and Vaccination. *Can J Infect Dis Med Microbiol.* 2023;20234263309. doi:10.1155/2023/4263309 Cited in: PubMed; PMID 36644336.
 15. Wang X, Ren J, Gao Q, Hu Z, Sun Y, Li X, Rowlands DJ, Yin W, Wang J, Stuart DI, Rao Z, Fry EE. Hepatitis A virus and the origins of picornaviruses. *Nature.* 2015;517(7532):85–8. doi:10.1038/nature13806 Cited in: PubMed; PMID 25327248.
 16. Hussain Z, Husain SA, Almajhdi FN, Kar P. Immunological and molecular epidemiological characteristics of acute and fulminant viral hepatitis A. *Viol J.* 2011;8254. doi:10.1186/1743-422X-8-254 Cited in: PubMed; PMID 21605420.
 17. Desbois D, Couturier E, Mackiewicz V, Graube A, Letort M-J, Dussaix E, Roque-Afonso A-M. Epidemiology and genetic characterization of hepatitis A virus genotype IIA. *J Clin Microbiol.* 2010;48(9):3306–15. doi:10.1128/JCM.00667-10 Cited in: PubMed; PMID 20592136.
 18. Nainan OV, Xia G, Vaughan G, Margolis HS. Diagnosis of hepatitis a virus infection: a molecular approach. *Clin Microbiol Rev.* 2006;19(1):63–79. doi:10.1128/CMR.19.1.63-79.2006 Cited in: PubMed; PMID 16418523.
 19. Li P, Liu J, Li Y, Su J, Ma Z, Bramer WM, Cao W, Man RA de, Peppelenbosch MP, Pan Q. The global epidemiology of hepatitis E virus infection: A systematic review and meta-analysis. *Liver International.* 2020;40(7):1516–28. doi:10.1111/liv.14468 Cited in: PubMed; PMID 32281721.
 20. Kamar N, Bendall R, Legrand-Abravanel F, Xia N-S, Ijaz S, Izopet J, Dalton HR. Hepatitis E. *Lancet.* 2012;379(9835):2477–88. doi:10.1016/S0140-6736(11)61849-7 Cited in: PubMed; PMID 22549046.

21. Smith DB, Simmonds P, Members OTICOTTOVSG, Jameel S, Emerson SU, Harrison TJ, Meng X-J, Okamoto H, van der Poel WHM, Purdy MA. Consensus proposals for classification of the family Hepeviridae. *J Gen Virol*. 2014;95(Pt 10):2223–32. doi:10.1099/vir.0.068429-0 Cited in: PubMed; PMID 24989172.
22. Khuroo MS. Study of an epidemic of non-A, non-B hepatitis. Possibility of another human hepatitis virus distinct from post-transfusion non-A, non-B type. *Am J Med*. 1980;68(6):818–24. doi:10.1016/0002-9343(80)90200-4 Cited in: PubMed; PMID 6770682.
23. Balayan MS, Andjaparidze AG, Savinskaya SS, Ketiladze ES, Braginsky DM, Savinov AP, Poleschuk VF. Evidence for a virus in non-A, non-B hepatitis transmitted via the fecal-oral route. *Intervirology*. 1983;20(1):23–31. doi:10.1159/000149370 Cited in: PubMed; PMID 6409836.
24. Teshale EH, Hu DJ, Holmberg SD. The two faces of hepatitis E virus. *Clin Infect Dis*. 2010;51(3):328–34. doi:10.1086/653943 Cited in: PubMed; PMID 20572761.
25. Carratalà A, Joost S. Population density and water balance influence the global occurrence of hepatitis E epidemics. *Sci Rep*. 2019;9(1):10042. doi:10.1038/s41598-019-46475-3 Cited in: PubMed; PMID 31296895.
26. Schwartz E, Jenks NP, van Damme P, Galun E. Hepatitis E virus infection in travelers. *Clin Infect Dis*. 1999;29(5):1312–4. doi:10.1086/313430 Cited in: PubMed; PMID 10524982.
27. Cordes AK, Goudeva L, Lütgehetmann M, Wenzel JJ, Behrendt P, Wedemeyer H, Heim A. Risk of transfusion-transmitted hepatitis E virus infection from pool-tested platelets and plasma. *J Hepatol*. 2022;76(1):46–52. doi:10.1016/j.jhep.2021.08.018 Cited in: PubMed; PMID 34461207.
28. Navaneethan U, Al Mohajer M, Shata MT. Hepatitis E and pregnancy: understanding the pathogenesis. *Liver International*. 2008;28(9):1190–9. doi:10.1111/j.1478-3231.2008.01840.x Cited in: PubMed; PMID 18662274.
29. Aziz AB, Øverbø J, Dudman S, Julin CH, Kwon YJG, Jahan Y, Ali M, Dembinski JL. Hepatitis E Virus (HEV) Synopsis: General Aspects and Focus on Bangladesh. *Viruses*. 2022;15(1). doi:10.3390/v15010063 Cited in: PubMed; PMID 36680103.
30. Nagashima S, Takahashi M, Kobayashi T, Nishizawa T, Nishiyama T, Primadharsini PP, Okamoto H. Characterization of the Quasi-Enveloped Hepatitis E Virus Particles Released by the Cellular Exosomal Pathway. *J Virol*. 2017;91(22). doi:10.1128/JVI.00822-17 Cited in: PubMed; PMID 28878075.
31. Ahmad I, Holla RP, Jameel S. Molecular virology of hepatitis E virus. *Virus Research*. 2011;161(1):47–58. doi:10.1016/j.virusres.2011.02.011 Cited in: PubMed; PMID 21345356.
32. Aslan AT, Balaban HY. Hepatitis E virus: Epidemiology, diagnosis, clinical manifestations, and treatment. *World J Gastroenterol*. 2020;26(37):5543–60. doi:10.3748/wjg.v26.i37.5543 Cited in: PubMed; PMID 33071523.

33. Laugel E, Hartard C, Jeulin H, Berger S, Venard V, Bronowicki J-P, Schvoerer E. Full-length genome sequencing of RNA viruses-How the approach can enlighten us on hepatitis C and hepatitis E viruses. *Reviews in medical virology*. 2021;31(4):e2197. doi:10.1002/rmv.2197 Cited in: PubMed; PMID 34260779.
34. Smith DB, Izopet J, Nicot F, Simmonds P, Jameel S, Meng X-J, Norder H, Okamoto H, van der Poel WHM, Reuter G, Purdy MA. Update: proposed reference sequences for subtypes of hepatitis E virus (species Orthohepevirus A). *J Gen Virol*. 2020;101(7):692–8. doi:10.1099/jgv.0.001435 Cited in: PubMed; PMID 32469300.
35. Lee G-H, Tan B-H, Teo EC-Y, Lim S-G, Dan Y-Y, Wee A, Aw PPK, Zhu Y, Hibberd ML, Tan C-K, Purdy MA, Teo C-G. Chronic Infection With Camelid Hepatitis E Virus in a Liver Transplant Recipient Who Regularly Consumes Camel Meat and Milk. *Gastroenterology*. 2016;150(2):355-7.e3. doi:10.1053/j.gastro.2015.10.048 Cited in: PubMed; PMID 26551551.
36. Khuroo MS, Khuroo MS, Khuroo NS. Transmission of Hepatitis E Virus in Developing Countries. *Viruses*. 2016;8(9). doi:10.3390/v8090253 Cited in: PubMed; PMID 27657112.
37. Lopez-Lopez P, Rialde, Maria de los Angeles, Frias M, García-Bocanegra I, Brieva T, Caballero-Gomez J, Camacho A, Fernández-Molera V, Machuca I, Gomez-Villamandos JC, Rivero A, Rivero-Juarez A. Risk factors associated with hepatitis E virus in pigs from different production systems. *Veterinary Microbiology*. 2018;224:88–92. doi:10.1016/j.vetmic.2018.08.020 Cited in: PubMed; PMID 30269796.
38. Adlhoch C, Avellon A, Baylis SA, Ciccaglione AR, Couturier E, Sousa R de, Epštein J, Ethelberg S, Faber M, Fehér Á, Ijaz S, Lange H, Mandáková Z, Mellou K, Mozalevskis A, Rimhanen-Finne R, Rizzi V, Said B, Sundqvist L, Thornton L, Tosti ME, van Pelt W, Aspinall E, Domanovic D, Severi E, Takkinen J, Dalton HR. Hepatitis E virus: Assessment of the epidemiological situation in humans in Europe, 2014/15. *J Clin Virol*. 2016;82:9–16. doi:10.1016/j.jcv.2016.06.010 Cited in: PubMed; PMID 27393938.
39. Santos-Silva S, López-López P, Gonçalves HMR, Rivero-Juarez A, van der Poel WHM, Nascimento MSJ, Mesquita JR. A Systematic Review and Meta-Analysis on Hepatitis E Virus Detection in Farmed Ruminants. *Pathogens*. 2023;12(4). doi:10.3390/pathogens12040550 Cited in: PubMed; PMID 37111437.
40. Schemmerer M, Wenzel JJ, Stark K, Faber M. Molecular epidemiology and genotype-specific disease severity of hepatitis E virus infections in Germany, 2010-2019. *Emerg Microbes Infect*. 2022;11(1):1754–63. doi:10.1080/22221751.2022.2091479 Cited in: PubMed; PMID 35713010.
41. Quer J, Colomer-Castell S, Campos C, Andrés C, Piñana M, Cortese MF, González-Sánchez A, García-Cehic D, Ibáñez M, Pumarola T, Rodríguez-Frías F, Antón A, Tabernero D. Next-Generation Sequencing for Confronting Virus

- Pandemics. *Viruses*. 2022;14(3). doi:10.3390/v14030600 Cited in: PubMed; PMID 35337007.
42. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: An overview. *Hum Immunol*. 2021;82(11):801–11. doi:10.1016/j.humimm.2021.02.012 Cited in: PubMed; PMID 33745759.
 43. Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. *Science*. 2003;300(5617):286–90. doi:10.1126/science.1084564 Cited in: PubMed; PMID 12690187.
 44. Slatko BE, Gardner AF, Ausubel FM. Overview of Next-Generation Sequencing Technologies. *Curr Protoc Mol Biol*. 2018;122(1):e59. doi:10.1002/cpmb.59 Cited in: PubMed; PMID 29851291.
 45. Schwarze K, Buchanan J, Fermont JM, Dreau H, Tilley MW, Taylor JM, Antoniou P, Knight SJL, Camps C, Pentony MM, Kvikstad EM, Harris S, Popitsch N, Pagnamenta AT, Schuh A, Taylor JC, Wordsworth S. The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. *Genet Med*. 2020;22(1):85–94. doi:10.1038/s41436-019-0618-7 Cited in: PubMed; PMID 31358947.
 46. Kumar KR, Cowley MJ, Davis RL. Next-Generation Sequencing and Emerging Technologies. *Semin Thromb Hemost*. 2019;45(07):661–73. doi:10.1055/s-0039-1688446 Cited in: PubMed; PMID 31096307.
 47. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. *Genomics*. 2016;107(1):1–8. doi:10.1016/j.ygeno.2015.11.003 Cited in: PubMed; PMID 26554401.
 48. Lewis EK, Haaland WC, Nguyen F, Heller DA, Allen MJ, MacGregor RR, Berger CS, Willingham B, Burns LA, Scott GBI, Kittrell C, Johnson BR, Curl RF, Metzker ML. Color-blind fluorescence detection for four-color DNA sequencing. *Proc Natl Acad Sci U S A*. 2005;102(15):5346–51. doi:10.1073/pnas.0501606102 Cited in: PubMed; PMID 15800037.
 49. Ciuffreda L, Rodríguez-Pérez H, Flores C. Nanopore sequencing and its application to the study of microbial communities. *Comput Struct Biotechnol J*. 2021;191497–511. doi:10.1016/j.csbj.2021.02.020 Cited in: PubMed; PMID 33815688.
 50. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol*. 2020;21(1):30. doi:10.1186/s13059-020-1935-5 Cited in: PubMed; PMID 32033565.
 51. Pomerantz A, Peñafiel N, Arteaga A, Bustamante L, Pichardo F, Coloma LA, Barrio-Amorós CL, Salazar-Valenzuela D, Prost S. Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. *Gigascience*. 2018;7(4). doi:10.1093/gigascience/giy033 Cited in: PubMed; PMID 29617771.
 52. Prunerì G, Braud F de, Sapino A, Aglietta M, Vecchione A, Giusti R, Marchiò C, Scarpino S, Baggi A, Bonetti G, Franzini JM, Volpe M, Jommi C. Next-

- Generation Sequencing in Clinical Practice: Is It a Cost-Saving Alternative to a Single-Gene Testing Approach? *Pharmacoecon Open*. 2021;5(2):285–98. doi:10.1007/s41669-020-00249-0 Cited in: PubMed; PMID 33660227.
53. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*. 2014;30(22):3276–8. doi:10.1093/bioinformatics/btu531 Cited in: PubMed; PMID 25095880.
 54. Pearson WR. Finding Protein and Nucleotide Similarities with FASTA. *Curr Protoc Bioinformatics*. 2016;533.9.1-3.9.25. doi:10.1002/0471250953.bi0309s53 Cited in: PubMed; PMID 27010337.
 55. R Core Team. R: A language and environment for statistical computing. New York: Springer; 2008.
 56. Hadley Wickham. ggplot2: Elegant Graphics for Data Analysis. New York: Springer; 2016.
 57. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, Jesus JG de, Main BJ, Tan AL, Paul LM, Brackney DE, Grewal S, Gurfield N, van Rompay KKA, Isern S, Michael SF, Coffey LL, Loman NJ, Andersen KG. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol*. 2019;20(1):8. doi:10.1186/s13059-018-1618-7 Cited in: PubMed; PMID 30621750.
 58. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30(14):3059–66. doi:10.1093/nar/gkf436 Cited in: PubMed; PMID 12136088.
 59. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80. doi:10.1093/molbev/mst010 Cited in: PubMed; PMID 23329690.
 60. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094–100. doi:10.1093/bioinformatics/bty191 Cited in: PubMed; PMID 29750242.
 61. Li H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics*. 2021;37(23):4572–4. doi:10.1093/bioinformatics/btab705 Cited in: PubMed; PMID 34623391.
 62. Bonenfant Q, Noé L, Touzet H. Porechop_ABI: discovering unknown adapters in Oxford Nanopore Technology sequencing reads for downstream trimming. *Bioinform Adv*. 2023;3(1):vbac085. doi:10.1093/bioadv/vbac085 Cited in: PubMed; PMID 36698762.
 63. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–3. doi:10.1093/bioinformatics/btu033 Cited in: PubMed; PMID 24451623.
 64. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference.

- Bioinformatics. 2019;35(21):4453–5. doi:10.1093/bioinformatics/btz305 Cited in: PubMed; PMID 31070718.
65. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9. doi:10.1093/bioinformatics/btp352 Cited in: PubMed; PMID 19505943.
 66. Vennema H. Protocol Molecular detection and typing of VP1 region of Hepatitis A virus (HAV). 2014.
 67. Wenzel JJ, Preiss J, Schemmerer M, Huber B, Plentz A, Jilg W. Detection of hepatitis E virus (HEV) from porcine livers in Southeastern Germany and high sequence homology to human HEV isolates. *J Clin Virol*. 2011;52(1):50–4. doi:10.1016/j.jcv.2011.06.006 Cited in: PubMed; PMID 21742549.
 68. Boxman ILA, Jansen CCC, Hägele G, Zwartkruis-Nahuis A, Cremer J, Vennema H, Tijsma ASL. Porcine blood used as ingredient in meat productions may serve as a vehicle for hepatitis E virus transmission. *International Journal of Food Microbiology*. 2017;257:225–31. doi:10.1016/j.ijfoodmicro.2017.06.029 Cited in: PubMed; PMID 28692890.
 69. Oxford Nanopore Technologies. Rapid sequencing DNA V14 - barcoding (SQK-RBK114.24 or SQK-RBK114.96) (RBK_9176_v114_revM_27Nov2022).
 70. HAVNET | RIVM [Internet]. 2026 [updated 2026 Feb 25; cited 2026 Feb 25]. Available from: <https://www.rivm.nl/en/havnet>
 71. Fitzpatrick AH, Rupnik A, O'Shea H, Crispie F, Keaveney S, Cotter P. High Throughput Sequencing for the Detection and Characterization of RNA Viruses. *Front Microbiol*. 2021;12:621719. doi:10.3389/fmicb.2021.621719 Cited in: PubMed; PMID 33692767.
 72. Marten D. Transparente Datenbankunterstützung für Analysen auf Big Data.
 73. Edgar RC. Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat Commun*. 2022;13(1):6968. doi:10.1038/s41467-022-34630-w Cited in: PubMed; PMID 36379955.
 74. Riaz N, Leung P, Barton K, Smith MA, Carswell S, Bull R, Lloyd AR, Rodrigo C. Adaptation of Oxford Nanopore technology for hepatitis C whole genome sequencing and identification of within-host viral variants. *BMC Genomics*. 2021;22(1):148. doi:10.1186/s12864-021-07460-1 Cited in: PubMed; PMID 33653280.
 75. Wang J, Ke YH, Zhang Y, Huang KQ, Wang L, Shen XX, Dong XP, Xu WB, Ma XJ. Rapid and Accurate Sequencing of Enterovirus Genomes Using MinION Nanopore Sequencer. *Biomed Environ Sci*. 2017;30(10):718–26. doi:10.3967/bes2017.097 Cited in: PubMed; PMID 29122092.
 76. Chiara M, D'Erchia AM, Gissi C, Manzari C, Parisi A, Resta N, Zambelli F, Picardi E, Pavesi G, Horner DS, Pesole G. Next generation sequencing of SARS-CoV-2 genomes: challenges, applications and opportunities. *Brief*

- Bioinform. 2021;22(2):616–30. doi:10.1093/bib/bbaa297 Cited in: PubMed; PMID 33279989.
77. Visser M, Bester R, Burger JT, Maree HJ. Next-generation sequencing for virus detection: covering all the bases. *Viol J.* 2016;1385. doi:10.1186/s12985-016-0539-x Cited in: PubMed; PMID 27250973.
 78. Batista FM, Stapleton T, Lowther JA, Fonseca VG, Shaw R, Pond C, Walker DI, van Aerle R, Martinez-Urtaza J. Whole Genome Sequencing of Hepatitis A Virus Using a PCR-Free Single-Molecule Nanopore Sequencing Approach. *Front Microbiol.* 2020;11874. doi:10.3389/fmicb.2020.00874 Cited in: PubMed; PMID 32523561.
 79. Lee G-Y, Park K, Lee Y-S, Kim JH, Byun KS, Kim J, Kim W-K, Song J-W. Molecular diagnosis of patients with hepatitis A virus infection using amplicon-based nanopore sequencing. *PloS one.* 2023;18(7):e0288361. doi:10.1371/journal.pone.0288361 Cited in: PubMed; PMID 37437048.
 80. Zinter MS, Mayday MY, Ryckman KK, Jelliffe-Pawlowski LL, DeRisi JL. Towards precision quantification of contamination in metagenomic sequencing experiments. *Microbiome.* 2019;7(1):62. doi:10.1186/s40168-019-0678-6 Cited in: PubMed; PMID 30992055.
 81. Baylis SA, Adlhoch C, Childs L. An Evaluation of Hepatitis E Virus Molecular Typing Methods. *Clin Chem.* 2021;68(1):181–91. doi:10.1093/clinchem/hvab186 Cited in: PubMed; PMID 34969109.
 82. Liu W-C, Lin C-P, Cheng C-P, Ho C-H, Lan K-L, Cheng J-H, Yen C-J, Cheng P-N, Wu I-C, Li I-C, Chang BC-H, Tseng VS, Chiu Y-C, Chang T-T. Aligning to the sample-specific reference sequence to optimize the accuracy of next-generation sequencing analysis for hepatitis B virus. *Hepatology International.* 2016;10(1):147–57. doi:10.1007/s12072-015-9645-x Cited in: PubMed; PMID 26208819.
 83. Lapointe HR, Dong W, Dong WWY, Kirkby D, Woods C, Poon AFY, Howe AYM, Harrigan PR, Brumme CJ. Validation of a Genotype-Independent Hepatitis C Virus Near-Whole Genome Sequencing Assay. *Viruses.* 2021;13(9). doi:10.3390/v13091721 Cited in: PubMed; PMID 34578305.
 84. Zsichla L, Zeeb M, Fazekas D, Áy É, Müller D, Metzner KJ, Kouyos RD, Müller V. Comparative Evaluation of Open-Source Bioinformatics Pipelines for Full-Length Viral Genome Assembly. *Viruses.* 2024;16(12). doi:10.3390/v16121824 Cited in: PubMed; PMID 39772134.
 85. Costa-Mattioli M, Di Napoli A, Ferré V, Billaudel S, Perez-Bercoff R, Cristina J. Genetic variability of hepatitis A virus. *J Gen Virol.* 2003;84(Pt 12):3191–201. doi:10.1099/vir.0.19532-0 Cited in: PubMed; PMID 14645901.
 86. Cristina J, Costa-Mattioli M. Genetic variability and molecular evolution of hepatitis A virus. *Virus Research.* 2007;127(2):151–7. doi:10.1016/j.virusres.2007.01.005 Cited in: PubMed; PMID 17328982.
 87. Balayan MS. Natural hosts of hepatitis A virus. *Vaccine.* 1992;10 Suppl 1S27-31. doi:10.1016/0264-410x(92)90537-t Cited in: PubMed; PMID 1335654.

88. Vaughan G, Xia G, Forbi JC, Purdy MA, Rossi LMG, Spradling PR, Khudyakov YE. Genetic relatedness among hepatitis A virus strains associated with food-borne outbreaks. *PloS one*. 2013;8(11):e74546. doi:10.1371/journal.pone.0074546 Cited in: PubMed; PMID 24223112.
89. Smith DB, Simmonds P. Classification and Genomic Diversity of Enterically Transmitted Hepatitis Viruses. *Cold Spring Harb Perspect Med*. 2018;8(9). doi:10.1101/cshperspect.a031880 Cited in: PubMed; PMID 29530950.
90. Primadharsini PP, Nagashima S, Okamoto H. Genetic Variability and Evolution of Hepatitis E Virus. *Viruses*. 2019;11(5). doi:10.3390/v11050456 Cited in: PubMed; PMID 31109076.
91. Smith DB, Purdy MA, Simmonds P. Genetic variability and the classification of hepatitis E virus. *J Virol*. 2013;87(8):4161–9. doi:10.1128/JVI.02762-12 Cited in: PubMed; PMID 23388713.
92. Domingo E, García-Crespo C, Lobo-Vega R, Perales C. Mutation Rates, Mutation Frequencies, and Proofreading-Repair Activities in RNA Virus Genetics. *Viruses*. 2021;13(9). doi:10.3390/v13091882 Cited in: PubMed; PMID 34578463.
93. Meng X-J. Expanding Host Range and Cross-Species Infection of Hepatitis E Virus. *PLoS Pathog*. 2016;12(8):e1005695. doi:10.1371/journal.ppat.1005695 Cited in: PubMed; PMID 27490119.
94. Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. Inexpensive multiplexed library preparation for megabase-sized genomes. *PloS one*. 2015;10(5):e0128036. doi:10.1371/journal.pone.0128036 Cited in: PubMed; PMID 26000737.
95. Oxford Nanopore Technologies. Nanopore-only Microbial Isolate Sequencing Solution (NO-MISS) - Rapid Barcoding Kit V14 (SQK-RBK114.24 or SQK-RBK114.96) (ISO_9205_v114_revF_13Dec2024) [Internet]. 2023 [updated 2024 Dec 15; cited 2024 Dec 15]. Available from: <https://nanoporetech.com/document/no-miss-isolate-sequencing-rapid-barcoding-v14>
96. Gilpatrick T, Lee I, Graham JE, Raimondeau E, Bowen R, Heron A, Downs B, Sukumar S, Sedlazeck FJ, Timp W. Targeted nanopore sequencing with Cas9-guided adapter ligation. *Nat Biotechnol*. 2020;38(4):433–8. doi:10.1038/s41587-020-0407-5 Cited in: PubMed; PMID 32042167.
97. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB. Characterizing and measuring bias in sequence data. *Genome Biol*. 2013;14(5):R51. doi:10.1186/gb-2013-14-5-r51 Cited in: PubMed; PMID 23718773.
98. Santos R, Lee H, Williams A, Baffour-Kyei A, Lee S-H, Troakes C, Al-Chalabi A, Breen G, Iacoangeli A. Investigating the Performance of Oxford Nanopore Long-Read Sequencing with Respect to Illumina Microarrays and Short-Read Sequencing. *Int J Mol Sci*. 2025;26(10). doi:10.3390/ijms26104492 Cited in: PubMed; PMID 40429637.

99. Maghini DG, Moss EL, Vance SE, Bhatt AS. Improved high-molecular-weight DNA extraction, nanopore sequencing and metagenomic assembly from the human gut microbiome. *Nat Protoc.* 2021;16(1):458–71. doi:10.1038/s41596-020-00424-x Cited in: PubMed; PMID 33277629.
100. Oxford Nanopore Technologies. Ligation sequencing amplicons V14 (SQK-LSK114) (ACDE_9163_v114_revV_12Dec2024) [Internet]. 2022 [updated 2024 Dec 13; cited 2024 Dec 13]. Available from: <https://nanoporetech.com/document/ligation-sequencing-amplicons-sqk-lsk114#issues-during-dna-rna-extraction-and-library-preparation>

9 Danksagung

„Wenn ich weitergesehen habe als andere, so deshalb, weil ich auf den Schultern von Riesen stand.“

(Isaac Newton)

Dieses Bild beschreibt treffend den Entstehungsprozess der vorliegenden Arbeit. Mein besonderer Dank gilt daher meinen persönlichen „Riesen“: Mathias Schemmerer, Jürgen Wenzel und Monika Erl, die mich auf diesem Weg gestützt und emporgehoben haben.

Mein großer Dank gilt meinem Doktorvater Prof. Dr. Jürgen Wenzel. Ich danke ihm herzlich für die Möglichkeit, in seiner Arbeitsgruppe zu promovieren, sowie für seine stetige Hilfsbereitschaft und das inspirierende, freundliche Umfeld, das er geschaffen hat.

Ein ebenso großer Dank gilt meinem Betreuer Dr. Mathias Schemmerer. Er hat mir nicht nur neue fachliche Ausblicke ermöglicht, sondern mit seiner unermüdlichen Geduld und fachlichen Exzellenz durch alle kleinen und großen Herausforderungen dieser Forschung geleitet.

Ebenso danke ich Monika Erl, die meine Arbeit mit ihrer großen Hilfsbereitschaft und ihrer herzlichen Art bereichert und das Labor mit Freude gefüllt hat.

Darüber hinaus gilt mein Dank dem gesamten Team der PCR-Abteilung am IMHR für die Bereitstellung der Proben und die reibungslose Zusammenarbeit.

Abschließend danke ich meiner Familie, die mir diesen Bildungsweg überhaupt erst ermöglicht hat und meinen Freunden für die Unterstützung und die vielen wertvollen Momente der Ablenkung.

10 Eidesstattliche Erklärung

Erklärung zur Anregung und Betreuung der Arbeit

Die vorliegende Dissertation wurde von Herrn Prof. Dr. Jürgen Wenzel angeregt und unter seiner Leitung am Lehrstuhl für Medizinische Mikrobiologie und Hygiene der Universität Regensburg ausgearbeitet und überwacht.

Eidesstattliche Erklärung

Ich, Renner Annika, geboren am 19.11.2000 in Heidenheim, erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Insbesondere habe ich nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeit erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt. Des Weiteren erkläre ich, dass die KI-basierten Werkzeuge ChatGPT und Gemini unterstützend zur Softwareprogrammierung sowie sprachlichen Textoptimierung und Rechtschreibprüfung eingesetzt wurden. Ich bin als Autorin dieser Dissertation für den gesamten Inhalt allein verantwortlich und trage selbstständig dafür Sorge, dass meine Dissertation inhaltlich und formal korrekt ist, die Regeln der guten wissenschaftlichen Praxis geachtet wurden und alle verwendeten Hilfsmittel korrekt angegeben wurden. Für etwaige Fehler durch den Einsatz von KI bin ich vollumfänglich verantwortlich.

Giengen, den 29.03.2026