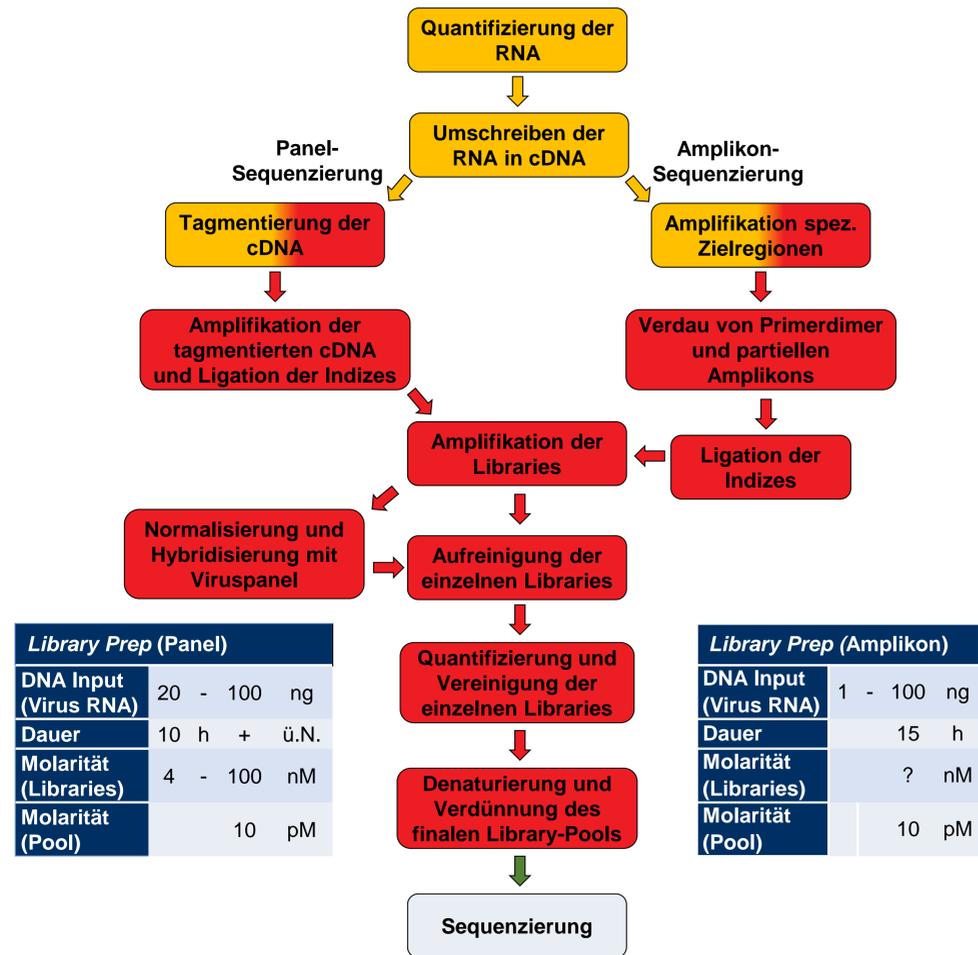


SARS-CoV-2-ANALYSE ZU VOC UND AUFKLÄRUNG VON INFEKTIONSKETTEN

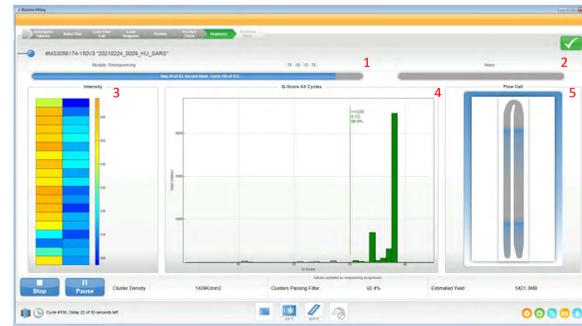
In der COVID-19 Pandemie unterstützt das NGS-Labor am Institut für Hygiene und Umwelt (NGS@HU) die Aufklärung von Infektionsketten und die Klassifizierung von SARS-CoV-2-Varianten, insbesondere von *variants of concern* (VOC). Für die Ganzgenomsequenzierung des 30 kb großen SARS-CoV-2-Virusgenoms wird *massive parallel sequencing* (MPS) auf einem Illumina MiSeq eingesetzt. Zur Herstellung der zu sequenzierenden Bibliotheken (*library prep*) sind zwei Methoden am HU in der Etablierung bzw. in der Anwendung. Dabei handelt es sich um die Panel- und Amplikon-Sequenzierung. Im Anschluss an die Sequenzierung wird mit rechnergestützten Analysen das Virusgenom anhand des Referenzgenoms (NC_045512) in seiner Gesamtheit rekonstruiert (*mapping, reference-based assembly*), und basierend auf den vorherrschenden Mutationen der Pangolin-Nomenklatur zugeordnet. Im Einzelnen werden die ermittelten DNA-Polymorphismen (*single nucleotide polymorphism* (SNP), Insertionen, Deletionen, etc.) im *variant call format* (vcf) gespeichert und als Grundlage weiterer Untersuchungen eingesetzt. Eine abgeleitete Consensus-Sequenz wird genutzt, um die vorherrschende SARS-CoV-2-Variante zu charakterisieren oder um Infektionsketten detailliert nachzuverfolgen.

I - Erstellung von DNA-Bibliotheken (*library prep*)



Die Erstellung von Sequenzier-Bibliotheken (*Library prep*), auf Basis von RNA ist komplexer und zeitaufwendiger als mit DNA als Ausgangsmaterial. Ein wesentlicher Unterschied besteht darin, dass bevor die DNA-Bibliothek erstellt werden kann, die RNA zunächst mittels reverse Transkription in cDNA umgeschrieben werden muss. Für die Sequenzierung von viraler RNA gibt es aktuell 3 Varianten. Bei der **Shotgun Sequenzierung** wird Patientenmaterial entnommen, unter Zellkulturbedingungen kultiviert und vermehrt. Nach Kultivierung wird aus der Reinkultur die RNA isoliert und die humane rRNA entfernt. Im Anschluss erfolgt die reverse Transkription und die Adapterligation an die cDNA. DNA-Fragmente mit beidseitigen Adaptern werden mit PCR angereichert und sequenziert. Bei **Amplikon- und Panel-Sequenzierungen** werden Virus-Panel eingesetzt. Gesamt-RNA wird aus Patientenmaterial isoliert und in cDNA umgeschrieben. Die Zielsequenz wird mit spezifischen Oligonukleotiden und PCR angereichert (Amplikon) oder mittels Hybridisierung (Panel) gewonnen. Oligonukleotide werden abgetrennt und Sequenzadapter ligiert. Vor Sequenzierung erhöht eine 2. Amplifikation die Quantität der spezifischen Virus-DNA mit Adaptern.

II - Sequenzierung und Laufparameter (Illumina MiSeq™)



Sequenzierparameter zur SARS-CoV-2-Analytik zur Bestimmung von VOC und zur Aufklärung von Infektionsketten. Diese Übersicht zeigt den Fortschritt¹, die Ausfälle², die Clusterdichte³, den Anteil über Qualitätswert Q30⁴ und die Darstellung einer Fließzelle⁵.

III - Bioinformatik

1. ASSEMBLIERUNG und QUALITÄTS-CHECK via Aquamis pipeline^[1]

sample name	species	total length	# contigs >1000 bp	Q30 base fraction	coverage depth	fraction majority species
2736Q	SARS coronavirus	34167	1	100%	196	0.98
2736V	SARS coronavirus	36610	4	100%	2704	0.87
2736Z	SARS coronavirus	38151	4	100%	1387	0.94
3122V	SARS coronavirus	32426	1	100%	1748	0.98
3225Q	SARS coronavirus	31125	1	100%	1403	0.94
3225V	SARS coronavirus	32041	2	100%	2423	0.97
3225Z	SARS coronavirus	33302	1	100%	473	0.64
3226V	SARS coronavirus	35572	4	100%	2712	0.81
818Q	SARS coronavirus	29870	1	100%	1692	0.94
818Z	SARS coronavirus	37520	5	100%	1757	0.85

[1] https://gitlab.com/bfr_bioinformatics/AQUAMIS

2. VOC-ANALYSE UND AUFKLÄRUNG VON AUSBRUCHSGESCHEHEN

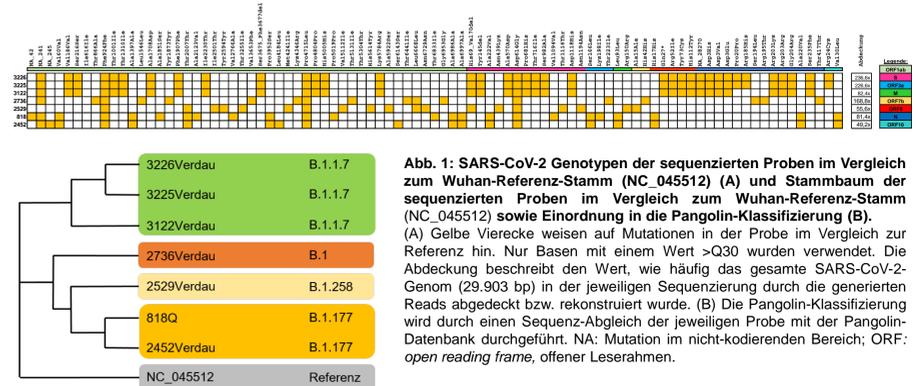


Abb. 1: SARS-CoV-2 Genotypen der sequenzierten Proben im Vergleich zum Wuhan-Referenz-Stamm (NC_045512) (A) und Stammbaum der sequenzierten Proben im Vergleich zum Wuhan-Referenz-Stamm (NC_045512) sowie Einordnung in die Pangolin-Klassifizierung (B). (A) Gelbe Vierecke weisen auf Mutationen in der Probe im Vergleich zur Referenz hin. Nur Basen mit einem Wert >Q30 wurden verwendet. Die Abdeckung beschreibt den Wert, wie häufig das gesamte SARS-CoV-2-Genom (29.903 bp) in der jeweiligen Sequenzierung durch die generierten Reads abgedeckt bzw. rekonstruiert wurde. (B) Die Pangolin-Klassifizierung wird durch einen Sequenz-Abgleich der jeweiligen Probe mit der Pangolin-Datenbank durchgeführt. NA: Mutation im nicht-kodierenden Bereich; ORF: open reading frame, offener Leserahmen.

Nach Qualitätskontrolle der *reads* werden die Virussequenzen basierend auf der SARS-CoV-2-Referenzgenomsequenz (NC045512.2) mittels *burrow wheeler aligner* (bwa) rekonstruiert (*mapping*). Mutationen (*single nucleotide polymorphisms* (SNP), Insertionen und Deletionen (Indels), etc.) innerhalb der Sequenzen werden mittels Genome Analysis Tool Kit (gatk) identifiziert und klassifiziert. Vorliegende VOC werden basierend auf der Pangolin-Klassifizierung eingeordnet. Eine Identifikation von Infektionsketten erfolgt abschließend durch R und Clusteranalysen.

NGS-LABOR (NGS@HU)

TYPISIERUNG VON SARS-CoV-2