

# Einfluss der Probenvorbereitung und Sequenzierung auf die Qualität von *short-read* *whole genome sequencing*-Daten

17.11.2023, Berlin

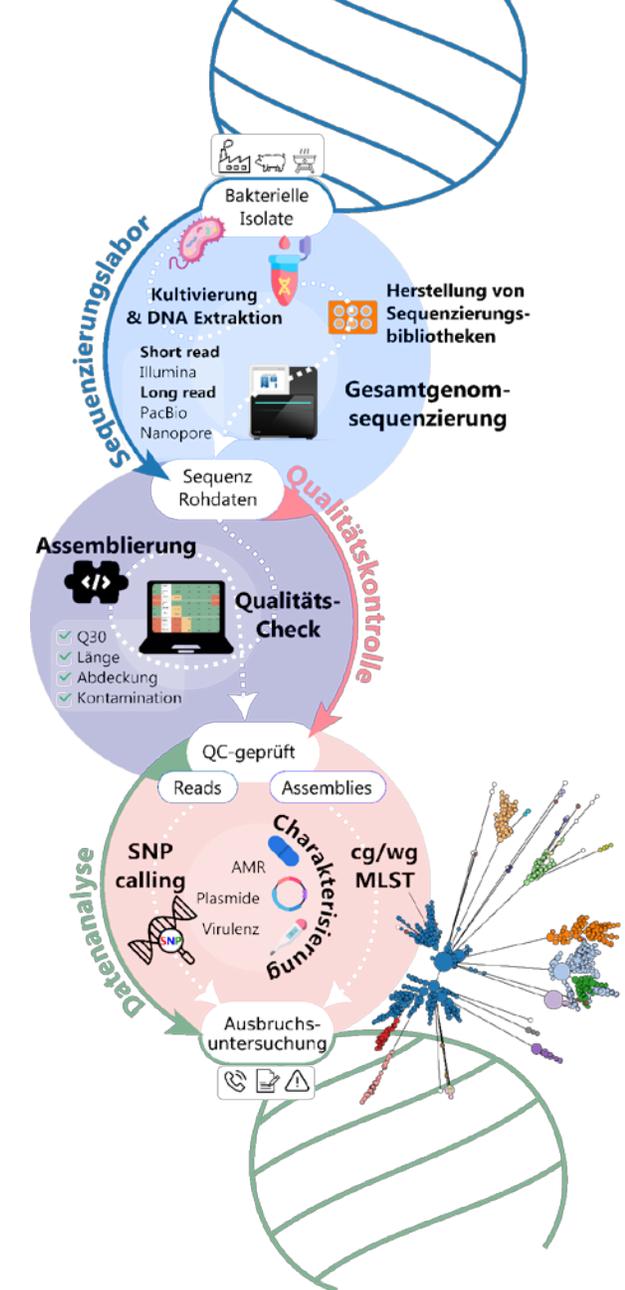
**Dr. Leonie Forth**

Nationales Studienzentrum für Sequenzierungen in der Risikobewertung

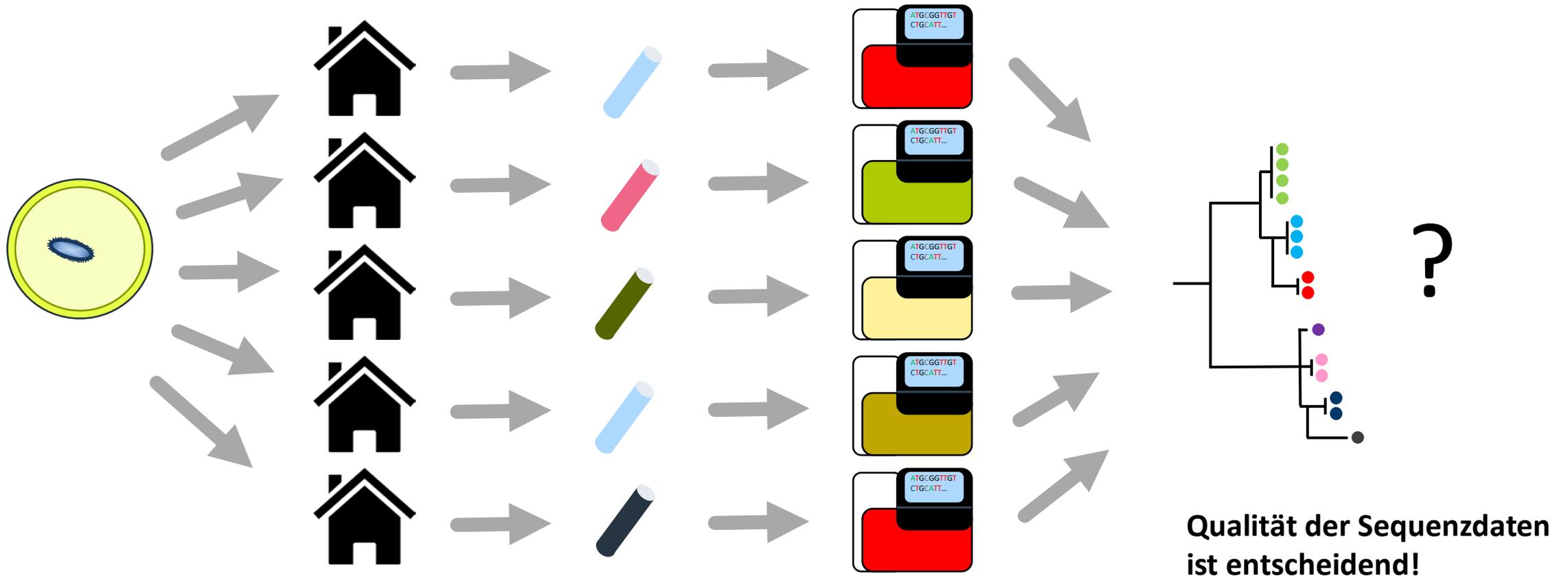
Abteilung Biologische Sicherheit

# Hintergrund

- Beteiligung **verschiedenster Labore** zur Generierung von bakteriellen Sequenzen, die in Ausbruchsuntersuchungen einbezogen werden
- **Keine einheitliche Vorgehensweise** zur Vorbereitung und Sequenzierung der Proben
- **Sicherstellung der Qualität** der erzeugten Sequenzen

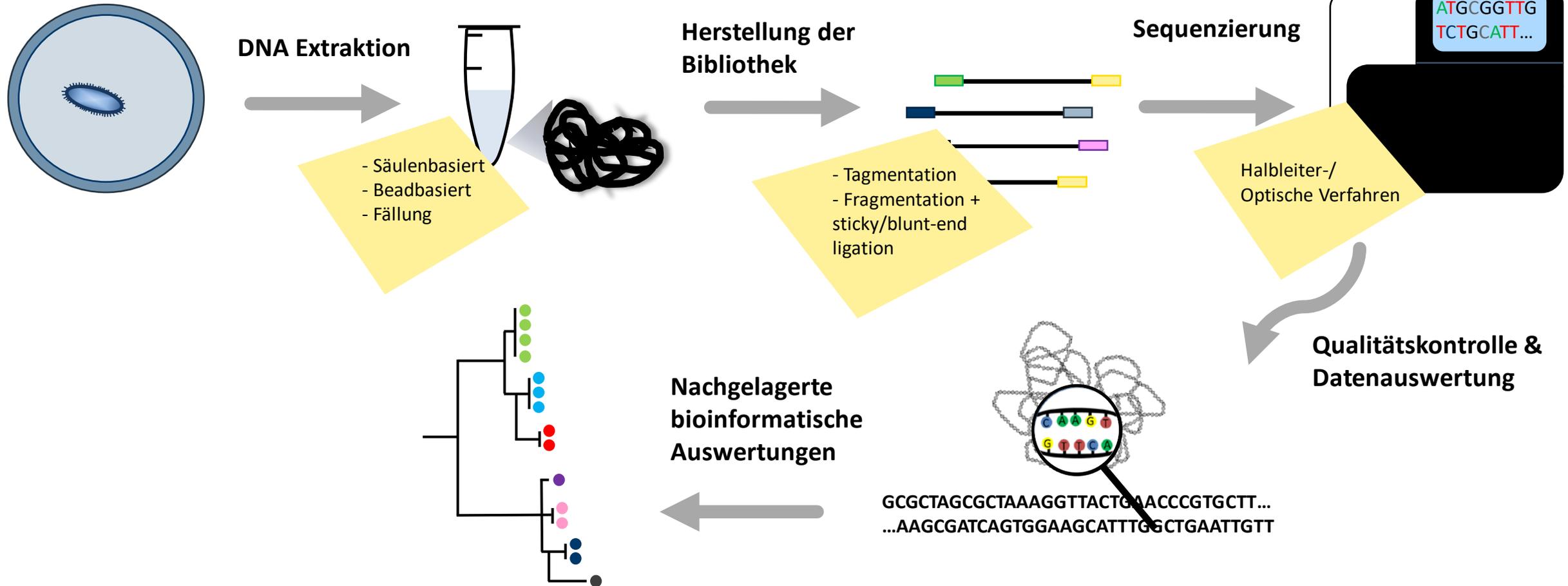


# Die Frage nach der Vergleichbarkeit



# Was kann die Qualität von WGS-Daten beeinflussen?

## - Ablauf der Probenvorbereitung und Sequenzierung -



# §64 LFGB AG „NGS-Bakteriencharakterisierung“



- Etabliert 2019 unter Leitung des BVL
- Landesuntersuchungsämter, Universitäten, Firmen, RKI, MRI, BVL, BfR
- Bearbeitung von NGS-Verfahren für die Sequenzierung bakterieller Erreger im Fokus auf *Listeria monocytogenes*, *Salmonella enterica*, STEC, *Campylobacter jejuni*
- **Methode** zur Gesamtgenomsequenzierung **entwickelt & veröffentlicht**
  - Beschreibung von Qualitätsmindestanforderungen
- Kurz vor Abschluss: **Leitlinie zur Clusteranalyse** von Gesamtgenomsequenzdaten
- Durchführung von Ringversuchen

März 2023

Amtliche Sammlung von Untersuchungsverfahren nach § 64 LFGB		
L	Untersuchung von Lebensmitteln Gesamtgenomsequenzierung zur Typisierung und Charakterisierung von <i>Salmonella enterica</i> , <i>Listeria monocytogenes</i> , <i>thermophilen Campylobacter spp.</i> und <b>Shigatoxin-bildenden</b> sowie <b>kommensalen <i>Escherichia coli</i></b> , die aus <b>Lebensmitteln</b> isoliert wurden	00.00
		183

# ISO 23418:2022

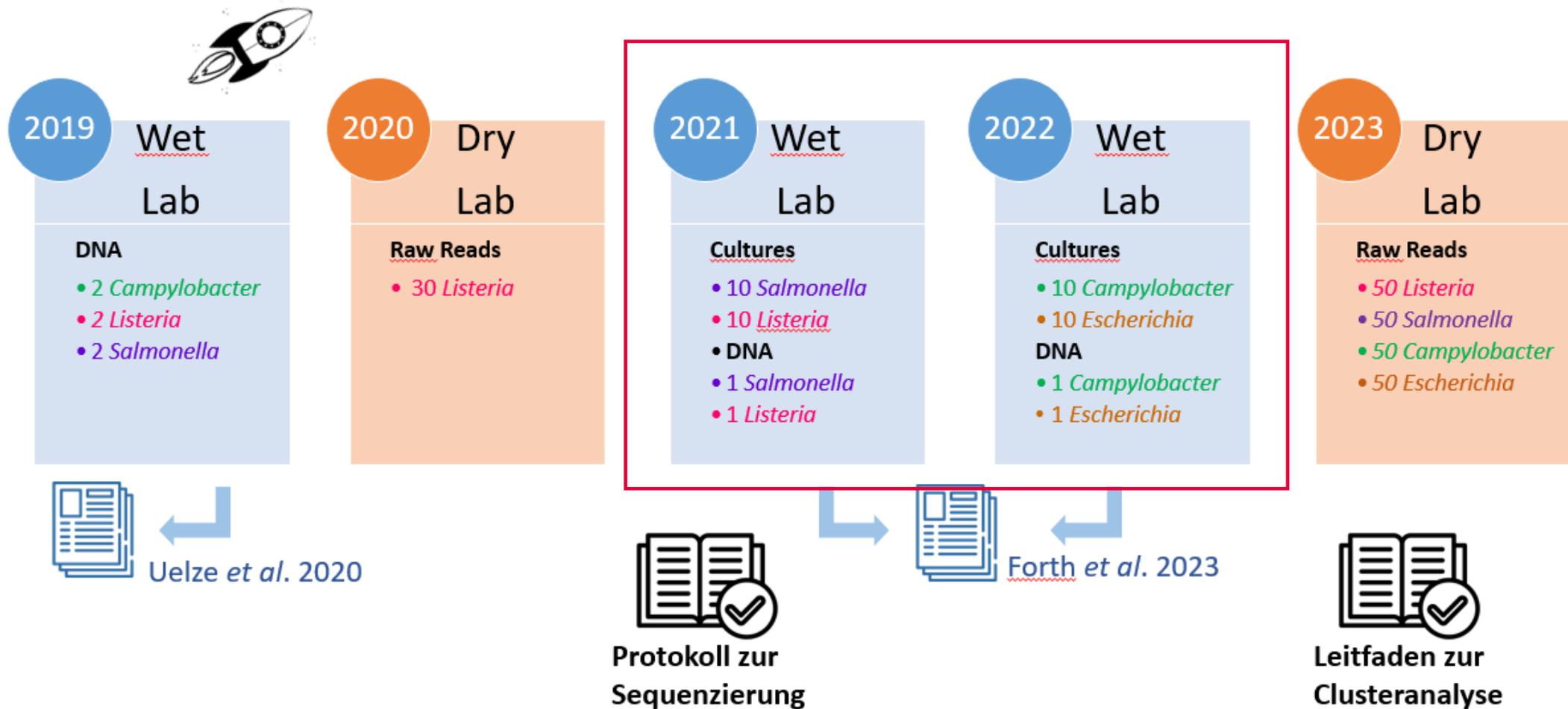
Spezifiziert Mindestanforderungen für die Erzeugung und Analyse von WGS Daten von Bakterien aus der Lebensmittelkette

DEUTSCHE NORM

September 2022

	<b>DIN EN ISO 23418</b>	<b><u>DIN</u></b>
ICS 07.100.30		
<b>Mikrobiologie der Lebensmittelkette - Gesamtgenomsequenzierung zur Typisierung und genomischen Charakterisierung von Bakterien - Allgemeine Anforderungen und Leitfaden (ISO 23418:2022); Deutsche Fassung EN ISO 23418:2022</b>		

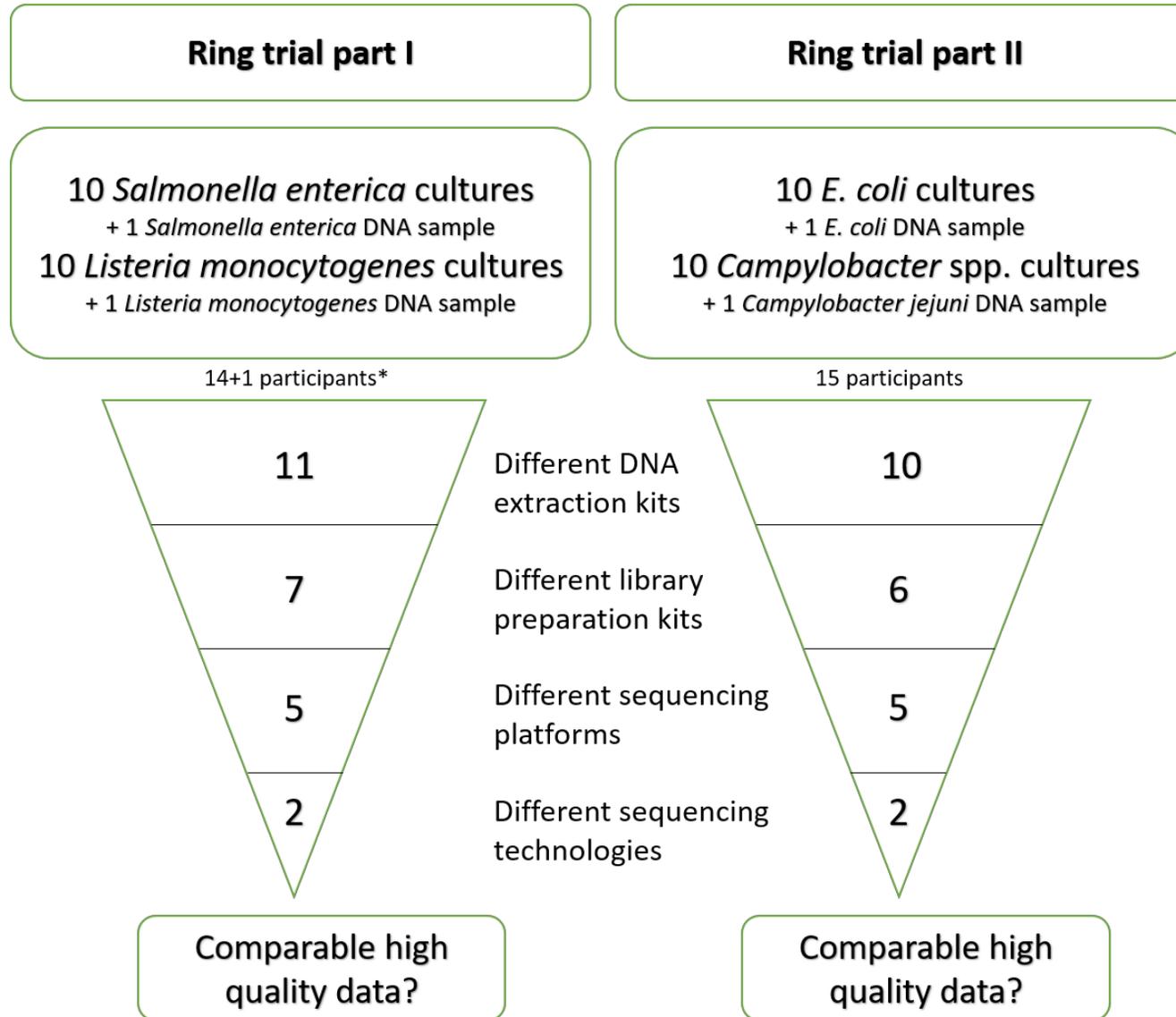
# Überblick der Ringversuche in der §64 LFGB Arbeitsgruppe „NGS-Bakteriencharakterisierung“



# Ringversuchsaufbau Wet-lab RV 2021 + 2022

- Je **10 Isolate als Kultur** und eine **DNA-Kontrolle** pro Spezies (+ Positiv- & Negativkontrolle)
- 2021: *L. monocytogenes* & *S. enterica*, 2022: *Campylobacter* spp. & STEC
- Je **14-15 teilnehmende Labore** (Landesuntersuchungsämter, Universitäten, Firmen, Bundesinstitute)
- Bearbeitungszeit: 12 Wochen
- Anwendung der **jeweiligen *in-house* Protokolle**
- **Fragebögen** zur Durchführung, Übermittlung von **Rohsequenzdaten**
- **Einheitliche Auswertung** durch 4NSZ mit BfR-eigenen open-source Pipelines

# Anwendung der *in-house* Protokolle

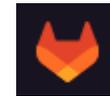


# Untersuchte Kriterien

- Fragebogen mit 52 Fragen zu: Lagerung, Anzucht, DNA-Extraktion, Bibliothekenherstellung, Sequenzierung, Auswertung
- Einheitliche Auswertung der Sequenzrohdaten mittels open-source Pipelines

659 Sequenzdatensätze

**AQUAMIS**, BakCharak, ChewieSnake, SnippySnake



[https://gitlab.com/bfr\\_bioinformatics](https://gitlab.com/bfr_bioinformatics)

## Levels of Analysis

- Participating laboratory
- Library preparation kit
- Sequencing platform
- Species
- Isolate

## Quality Assessment

- Q30 Base Fraction
- Assembly length
- Assembly coverage
- N50
- Full genes
- Contamination
- Assembler performance
- GC-Gehalt
- Abgeleitetes Genus

## Genetic Distance

- cgMLST distance
- SNP distance
- Sequence type

# Ziele der Untersuchung – Die großen Fragen

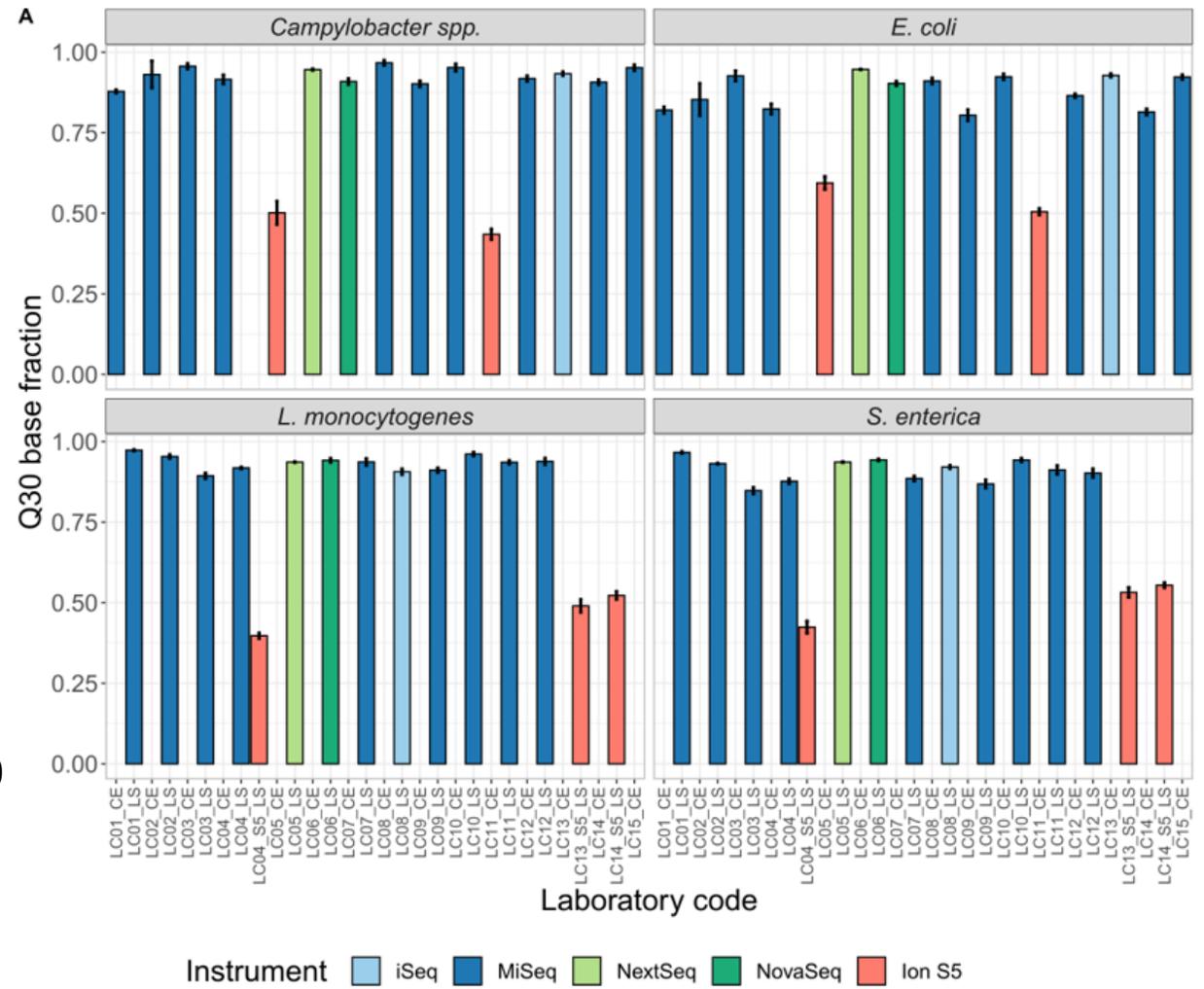


1. Wie groß sind die Unterschiede?
2. Wo werden die größten Unterschiede beobachtet?
3. Welche Ursache haben diese Unterschiede?
4. Welche Erkenntnisse können daraus abgeleitet werden?

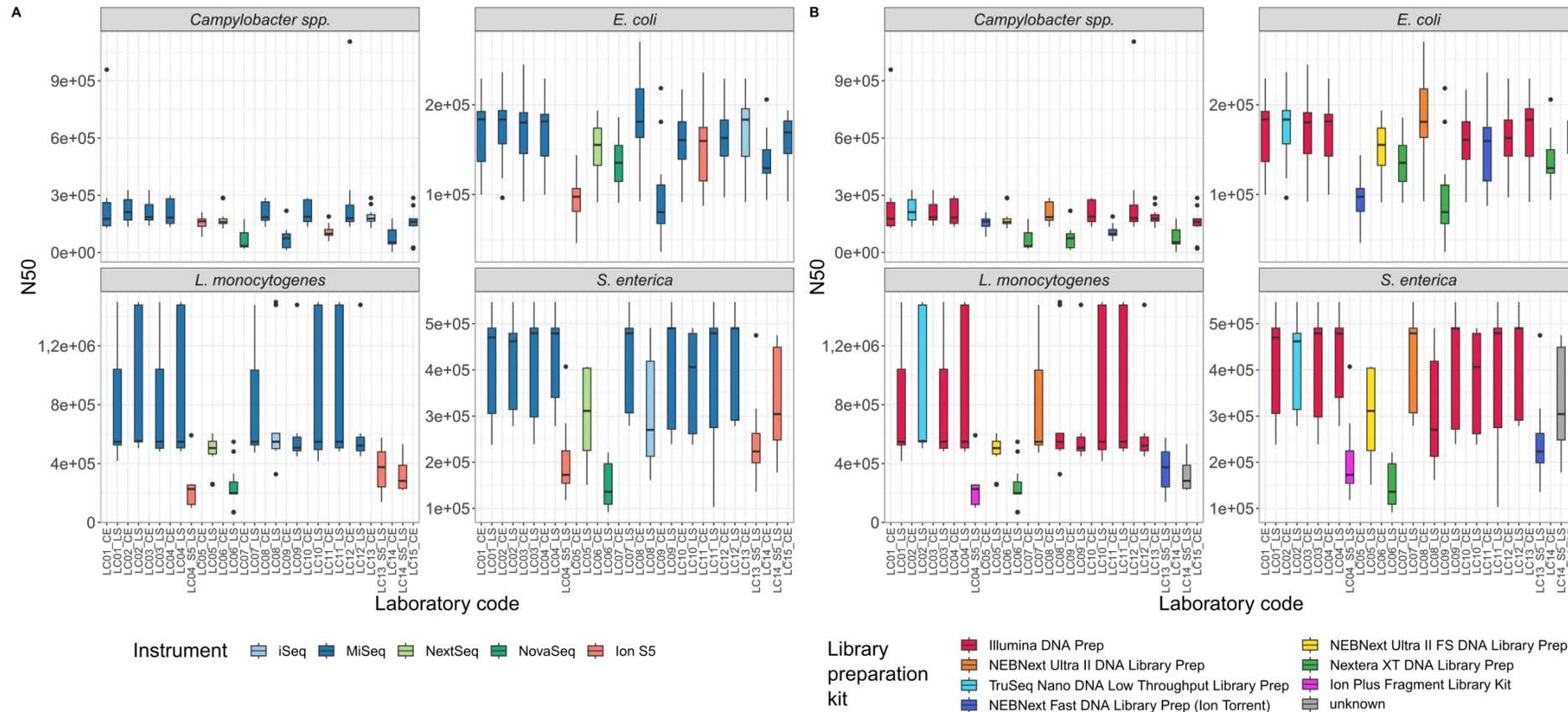
# Prozentualer Anteil an Basen mit einem Phred-Score $\geq 30$

Phred Quality Score	10	20	30
Probability of Incorrect Base Call	1 in 10	1 in 100	1 in 1000
Base Call Accuracy	90 %	99 %	99.9 %

Ion S5-Geräte generieren technologisch bedingt einen geringeren Anteil an Basen mit einem Phred-Score  $\geq 30$  als Illumina-Geräte.

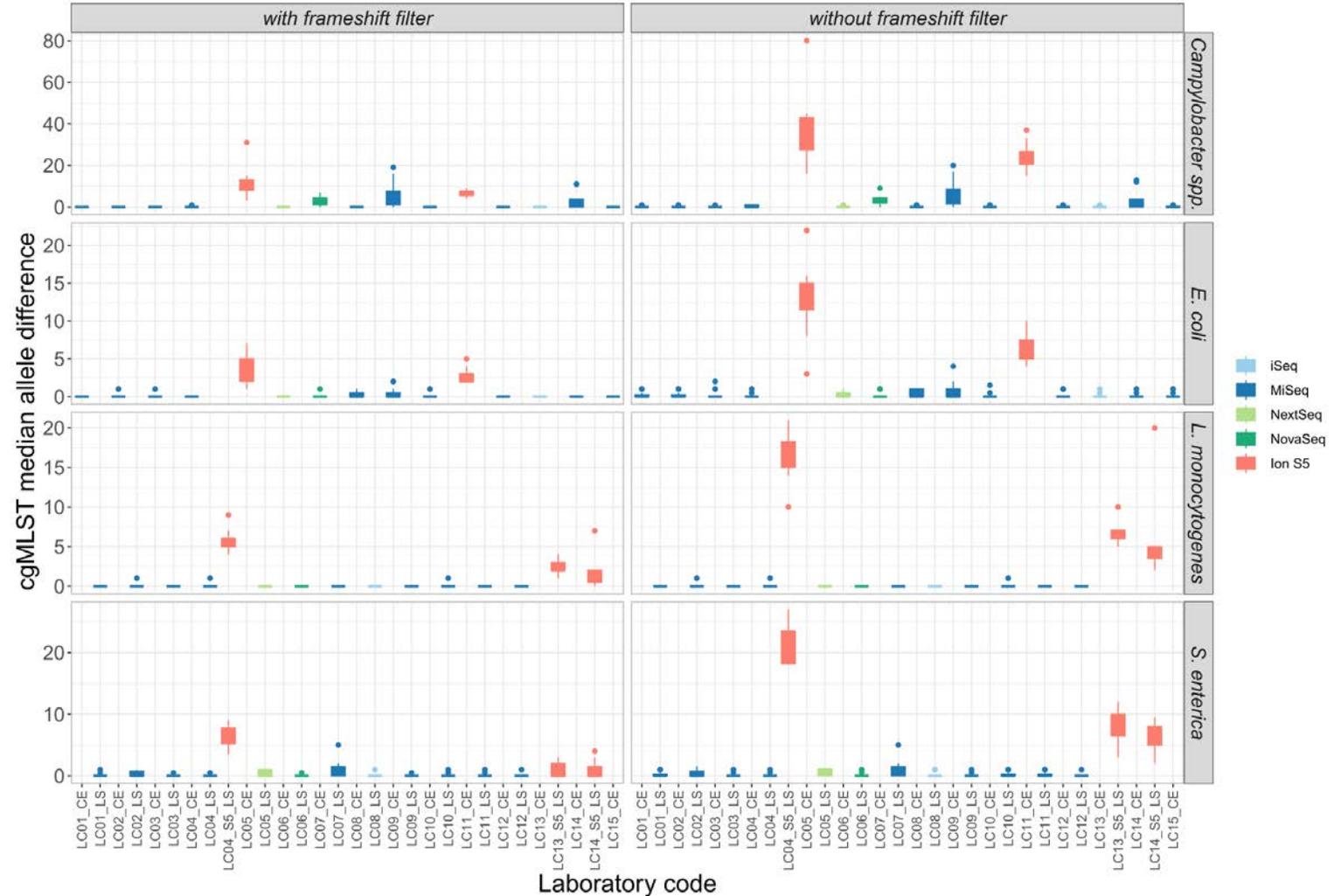


# N50-Wert

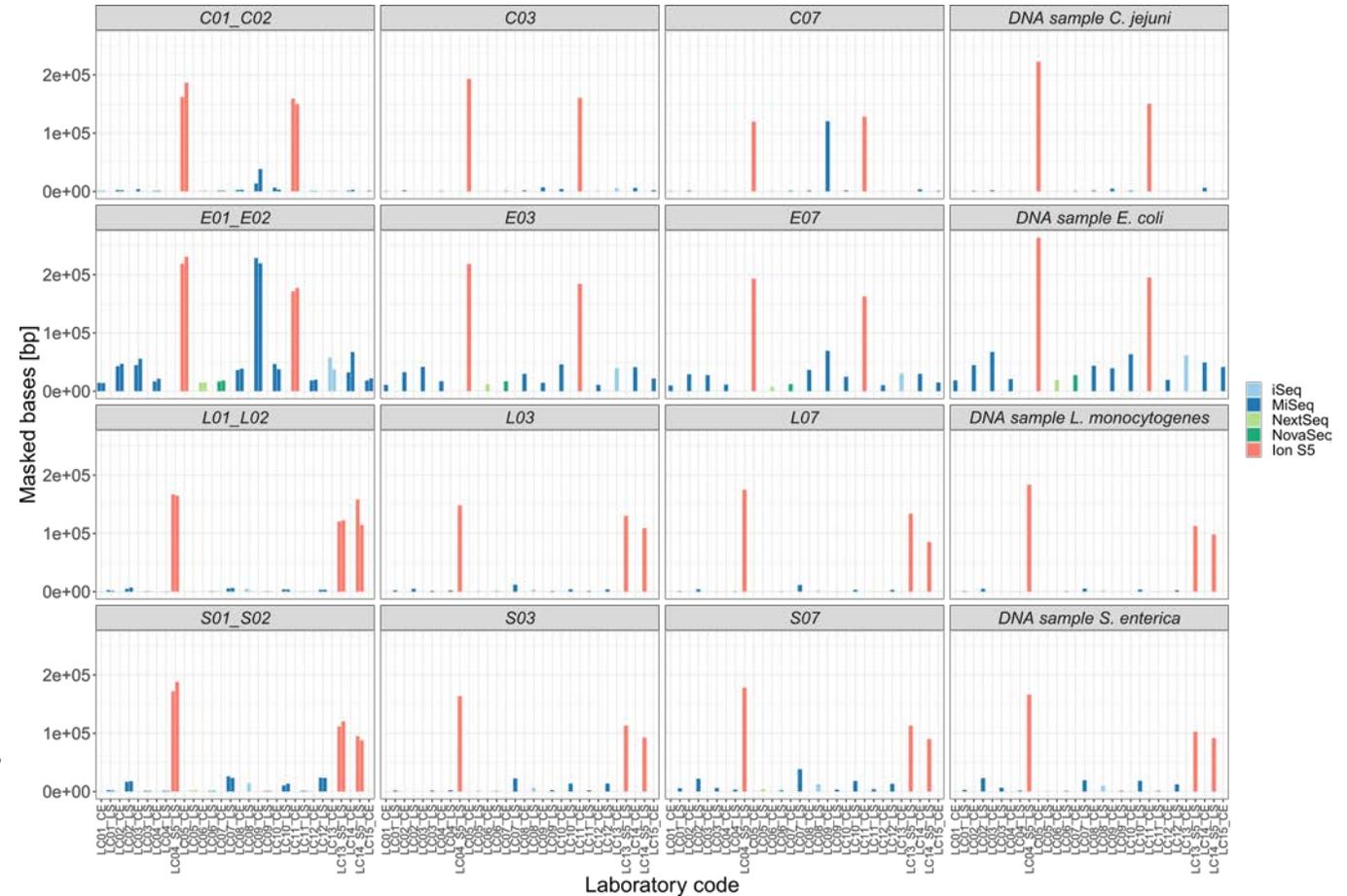
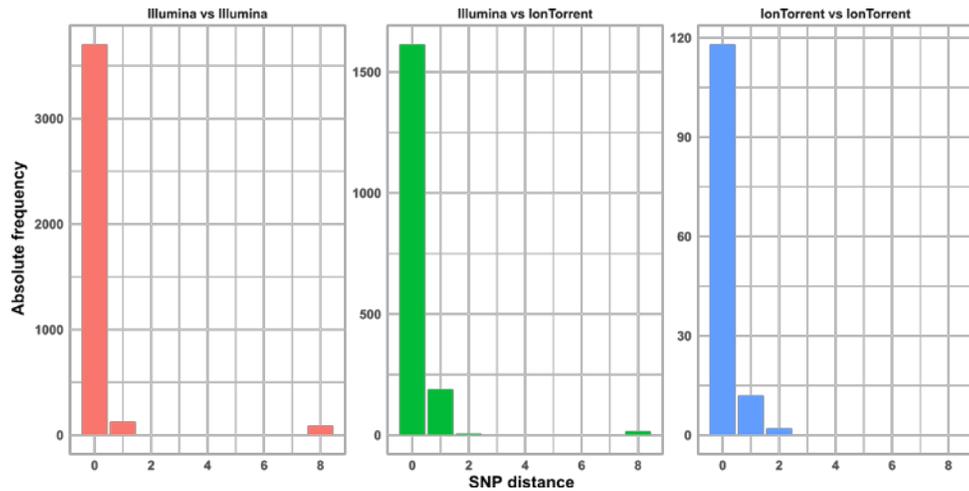


# cgMLST Analysen

Unterschiede im Median der Alleldifferenzen sind abhängig von der verwendeten Sequenziertechnologie als auch zum Teil vom verwendeten Kit zur Herstellung der Bibliothek.



# SNP Analysen



Die beobachteten SNP-Distanzen sind vergleichbar.  
Aber: Die Anzahl der maskierten Basen ist bei Ion S5-generierten Sequenzdaten deutlich erhöht!

# Vereinfachte Beantwortung der großen Fragen



## 1. Wie groß sind die Unterschiede?

Im Allgemeinen gering, mit deutlichen Ausnahmen

## 2. Wo werden die größten Unterschiede beobachtet?

Illumina <-> Ion Torrent; Nextera XT <-> Illumina DNA Prep Tagmentation; am deutlichsten sichtbar bei *Campylobacter*-Isolaten

## 3. Welche Ursache haben diese Unterschiede?

Unterschiedliche Sequenziertechnologie, GC-Bias des Library Prep Kits

## 4. Welche Erkenntnisse können daraus abgeleitet werden?

Vorsicht bei Analyse von Sequenzdaten unterschiedlicher Sequenziertechnologien (ggf. filtern),  
Vermeidung von Kits mit einem bekannten GC-Bias

# Detaillierte Beantwortung der Fragen



## Impact of wet-lab protocols on quality of whole-genome short-read sequences from foodborne microbial pathogens

1 Leonie F. Forth<sup>1</sup>, Erik Brinks<sup>2</sup>, Grégoire Denay<sup>3</sup>, Ahmad Fawzy<sup>4,5</sup>, Stefan Fiedler<sup>6</sup>, Jannika Fuchs<sup>7</sup>, Anne-Catrin Geuthner<sup>8</sup>, Thomas Hankeln<sup>9,10</sup>, Ekkehard Hiller<sup>11</sup>, Larissa Murr<sup>12</sup>, Henning Petersen<sup>13</sup>, Ralf Reiting<sup>14</sup>, Christian Schäfers<sup>15</sup>, Claudia Schwab<sup>16</sup>, Kathrin Szabo<sup>6</sup>, Andrea Thürmer<sup>17</sup>, Anne Wöhlke<sup>18</sup>, Jennie Fischer<sup>1</sup>, Stefanie Lüth<sup>1</sup>, Michaela Projahn<sup>1</sup>, Kerstin Stingl<sup>1</sup>, Maria Borowiak<sup>1</sup>, Carus Deneke<sup>1</sup>, Burkhard Malorny<sup>1\*</sup>, Laura Uelze<sup>1\*</sup>

Publikation akzeptiert zur Veröffentlichung in  
Frontiers in Microbiology,  
doi: 10.3389/fmicb.2023.1253362



Bioproject  
PRJEB62505



Detaillierte Reports der beiden Ringversuche 2021 und 2022 auf der Homepage des BVL

Vielen Dank für Ihre  
Aufmerksamkeit!

## Herzlichen Dank

dem gesamten Team des Nationalen  
Studienzentrums für  
Sequenzierungen in der  
Risikobewertung, u.a.

- Laura Uelze
- Burkhard Malorny
- Maria Borowiak
- Carlus Deneke
- Holger Brendebach

**& allen Teilnehmenden der  
Ringversuche!**



Dr. Leonie Forth  
T +49 30 18412-24914  
leonie.forth@bfr.bund.de

Bundesinstitut für Risikobewertung  
bfr.bund.de

**BfR** | Risiken erkennen –  
Gesundheit schützen

Verbraucherschutz zum Mitnehmen

**BfR2GO – das Wissenschaftsmagazin des BfR**

[bfr.bund.de/de/wissenschaftsmagazin\\_bfr2go.html](https://bfr.bund.de/de/wissenschaftsmagazin_bfr2go.html)

Folgen Sie uns

 @bfrde | @bfren | @Bf3R\_centre

 @bfrde

 youtube.com/@bfr\_bund

 social.bund.de/@bfr

 linkedin.com/company/bundesinstitut-f-r-risikobewertung