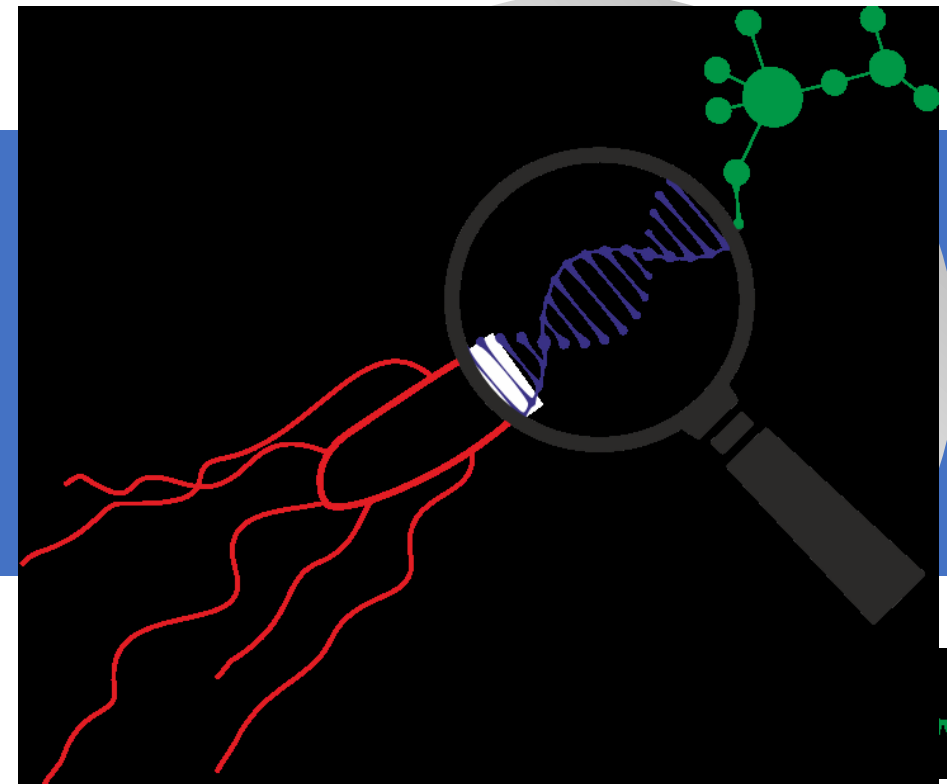


GenoSalmSurv screencasts



GenoSalmSurv screencasts

das Projekt

1

die Pipelines

AQUAMIS und die QC Entscheidung

Allele calling mit chewieSnake

Gemeinsame GenoSalmSurv Datenbank

Metachewiereport + Interpretation

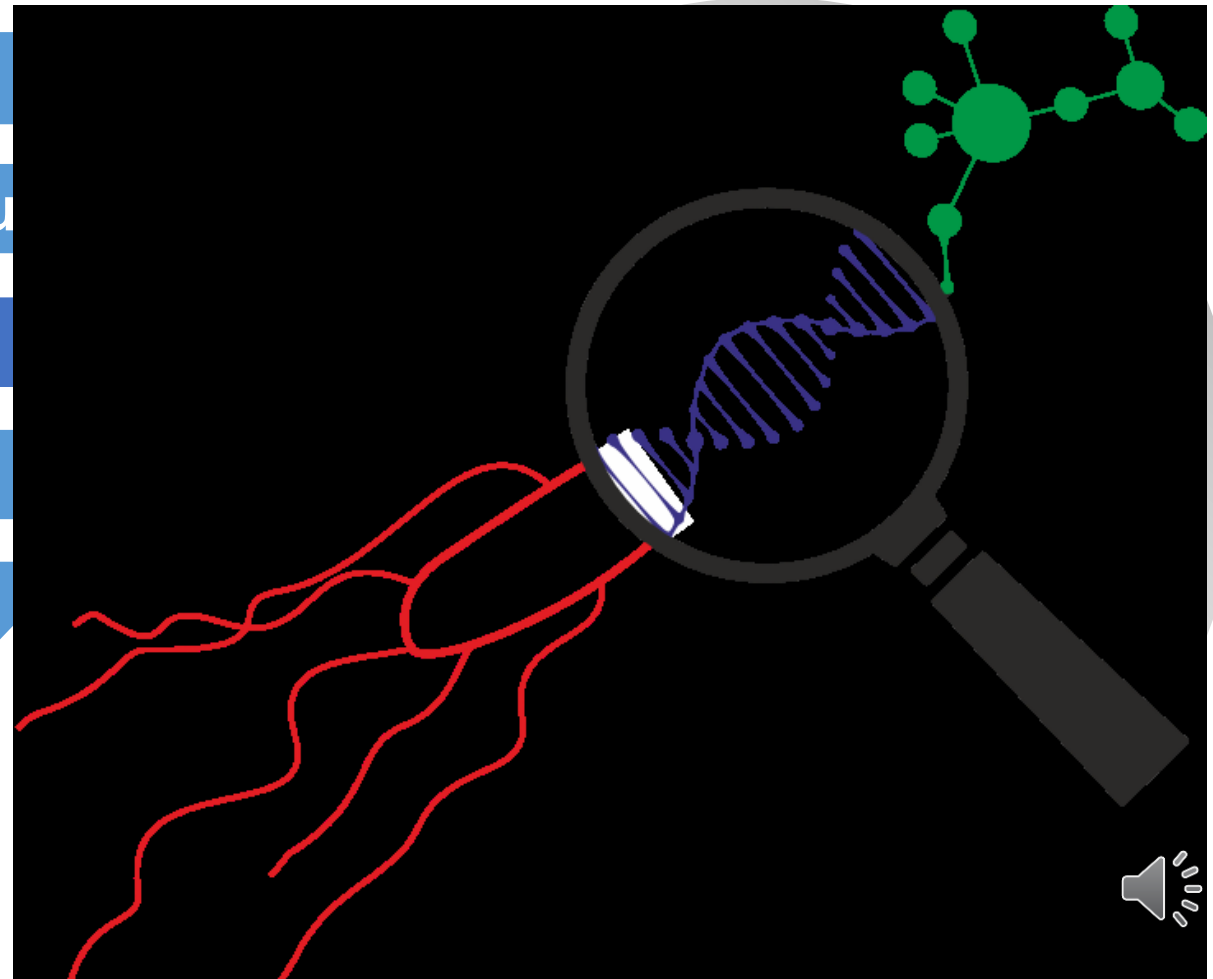
6

Demo einer Ausbruchsanalyse

7

Installationshilfe

8

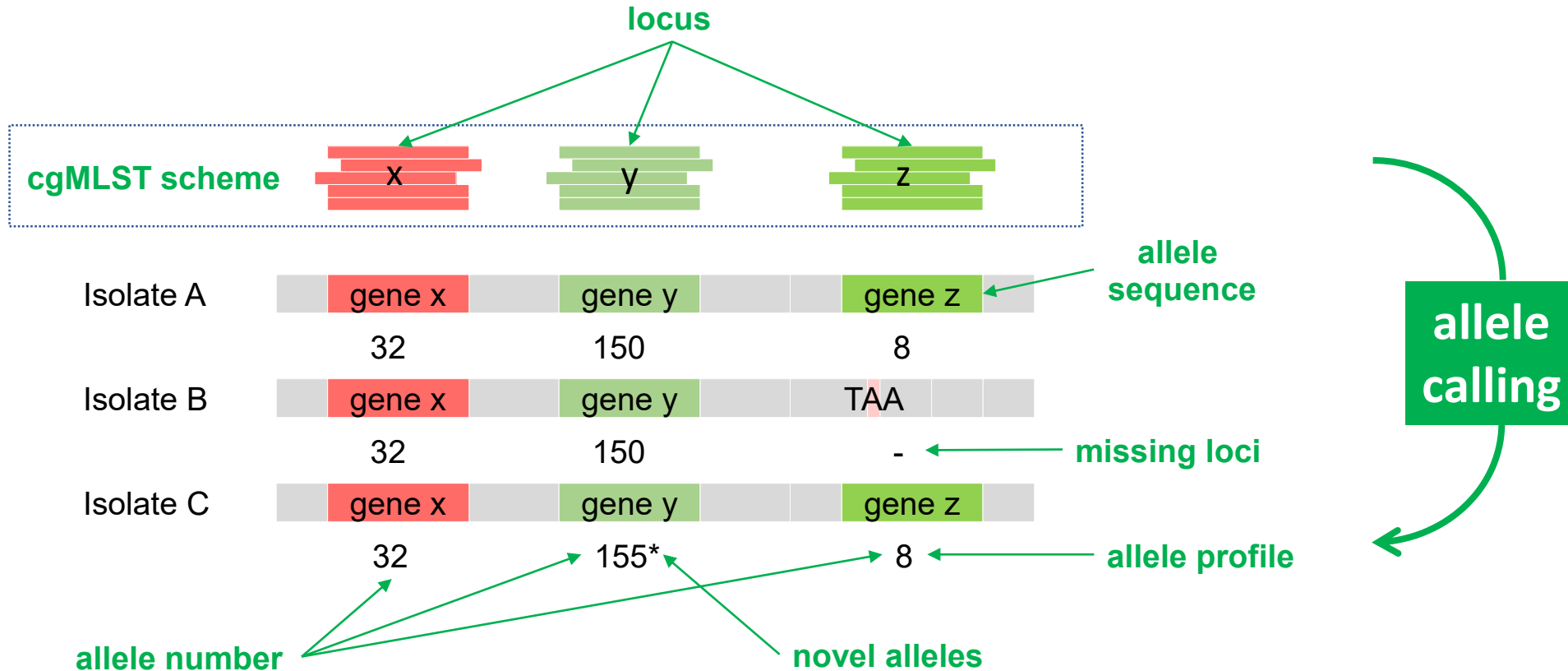


Inhalt

- Grundlagen und Begriffe cgMLST
- Überblick chewieSnake Pipeline
- Details zu Komponenten der chewieSnake Pipeline
 - Allele hashing
 - Clustering
 - Allele quality
 - Reporting und Visualisierung
 - Anwendung in der Routine durch Probenmanagement
- Zusammenfassung und weiterführende Informationen



Grundlagen und Begriffe cgMLST



Grundlagen und Begriffe cgMLST II

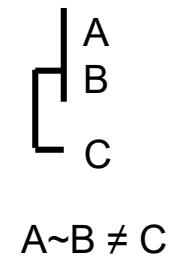
allele profile

	gene x	gene y	gene z
A	32	150	8
B	32	150	-
C	32	155	8

distance matrix

	A	B	C
A	-	0	1
B	0	-	1
C	1	1	-

phylogeny



Überblick chewieSnake Pipeline

Snakemake

Pipeline Manager zur Programmkontrolle und zum Ressourcenmanagement



Überblick chewieSnake Pipeline

Proben-
management

Sample sheet

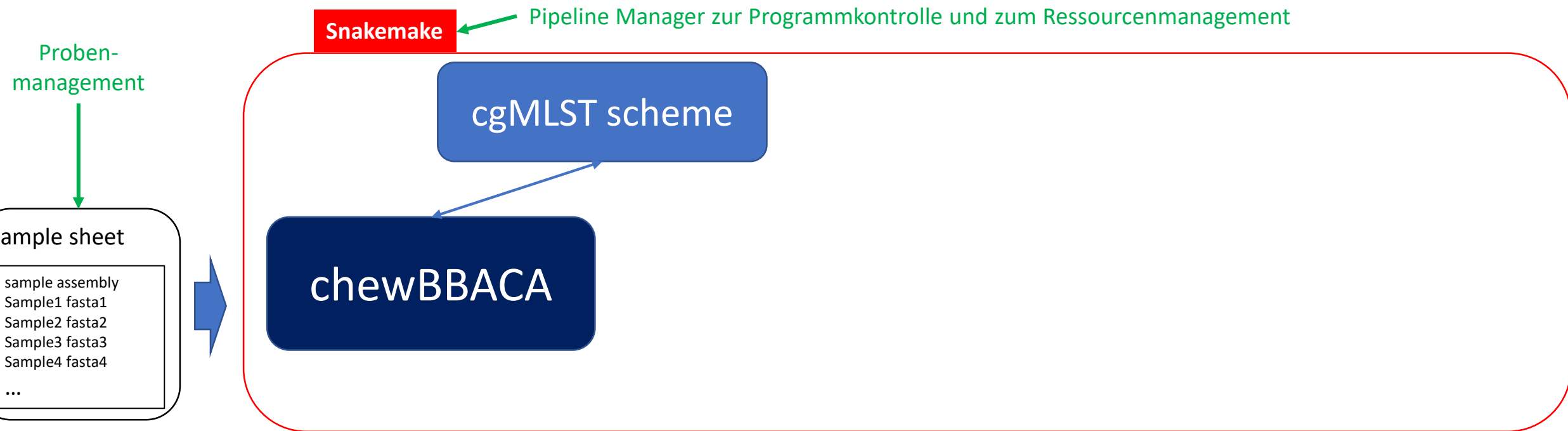
```
sample assembly
Sample1 fasta1
Sample2 fasta2
Sample3 fasta3
Sample4 fasta4
...
```

Snakemake

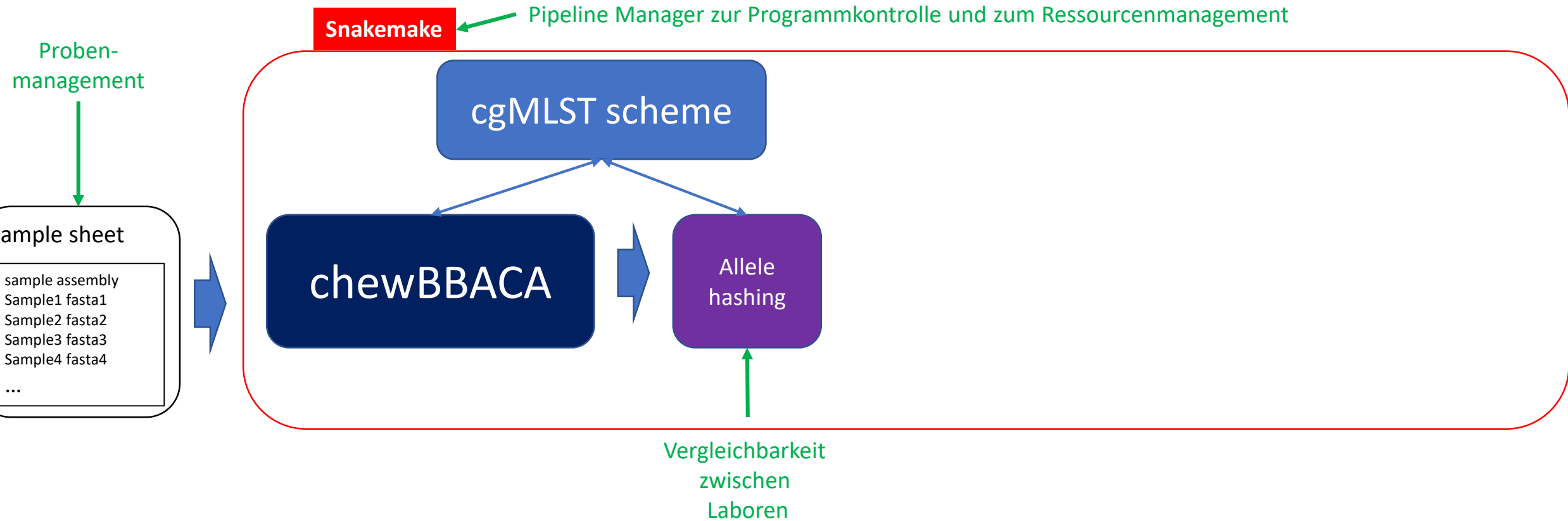
Pipeline Manager zur Programmkontrolle und zum Ressourcenmanagement



Überblick chewieSnake Pipeline



Überblick chewieSnake Pipeline



Überblick chewieSnake Pipeline

Snakemake

Pipeline Manager zur Programmkontrolle und zum Ressourcenmanagement

Proben-
management

Sample sheet

```
sample assembly
Sample1 fasta1
Sample2 fasta2
Sample3 fasta3
Sample4 fasta4
...
```

cgMLST scheme

chewBBACA

Allele
hashing

grapetree

Minimum
spanning
tree

Allele
distance
matrix

Vergleichbarkeit
zwischen
Laboren

Vergleichende
Analysen, Phylogenie



Überblick chewieSnake Pipeline

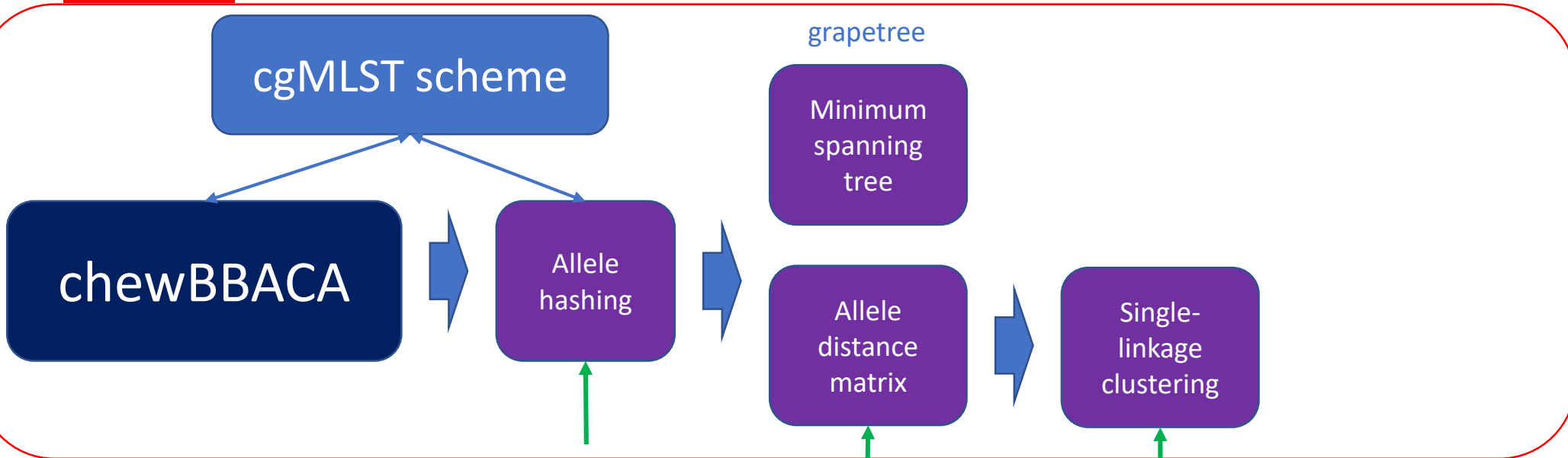
Snakemake

Pipeline Manager zur Programmkontrolle und zum Ressourcenmanagement

Proben-
management

Sample sheet

```
sample assembly  
Sample1 fasta1  
Sample2 fasta2  
Sample3 fasta3  
Sample4 fasta4  
...
```



Vergleichbarkeit
zwischen
Laboren

Vergleichende
Analysen, Phylogenie

Cluster
Nomenklatur



Überblick chewieSnake Pipeline

Snakemake

Pipeline Manager zur Programmkontrolle und zum Ressourcenmanagement

Proben-
management

Sample sheet

```
sample assembly
Sample1 fasta1
Sample2 fasta2
Sample3 fasta3
Sample4 fasta4
...
```



chewBBACA

cgMLST scheme

Allele
hashing

Vergleichbarkeit
zwischen
Laboren

grapetree

Minimum
spanning
tree

Allele
distance
matrix

Vergleichende
Analysen, Phylogenie

Single-
linkage
clustering

Cluster
Nomenklatur

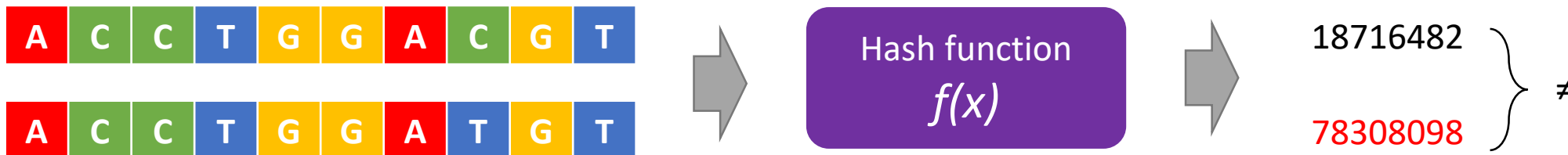
Ergebnis-
zusammen-
fassung

HTML
report



Allele hashing

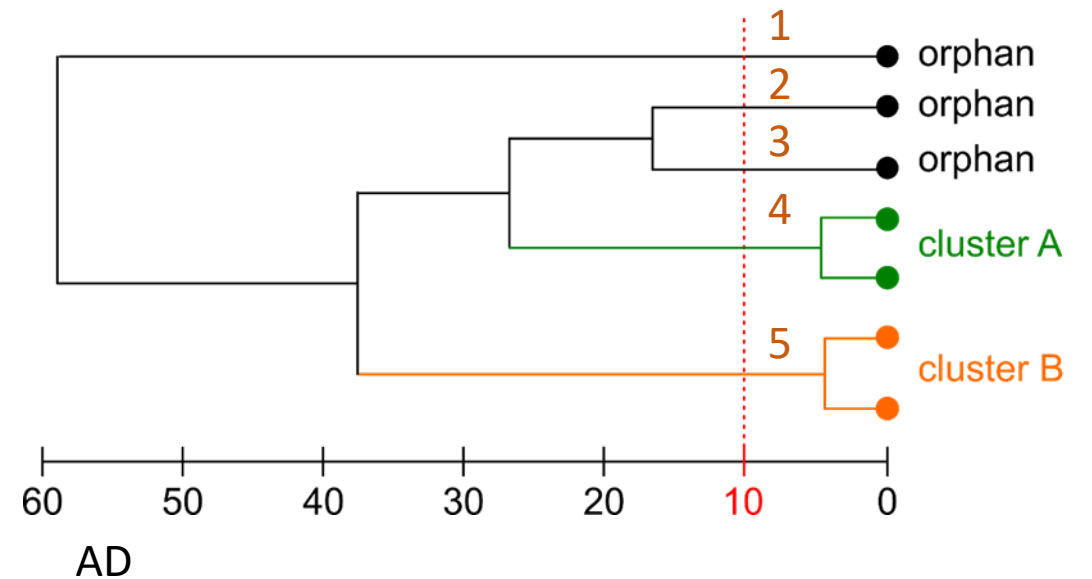
- Klassisch werden Allelsequenzen mit fortlaufenden Allelnummern bezeichnet
 - Neue Allelsequenzen bekommen eine neue Allelnummer
- Diese Benennung muss mittels eines zentralen Nomenklaturservers (NS) ausgetauscht werden, um vergleichbare Ergebnisse zu erzielen
 - Andernfalls können identische Allelsequenzen verschieden bezeichnet werden: Überschätzung AD
 - Verschiedene Allelsequenzen können identisch bezeichnet werden: Unterschätzung AD
- Allele hashing: Übersetzen einer Allelsequenz in einen eindeutigen Allele hash



Clustering der Proben

- Methode: Single-linkage hierarchical clustering der Distanzmatrix
- Benennung von Cluster-Typen zu einem Alleldistanz-Schwellenwert

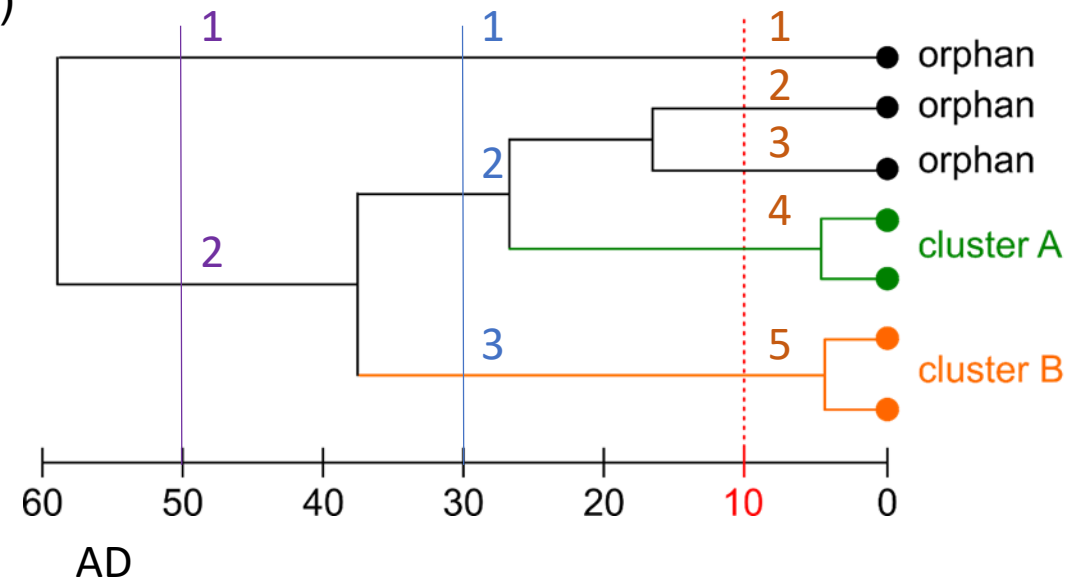
Probe			CT_10	
s1			1	
s2			2	
s3			3	
s4			4	
s5			4	
s6			5	
s7			5	



Clustering der Proben

- Methode: Single-linkage hierarchical clustering der Distanzmatrix
- Benennung von Cluster-Typen zu einem Alleldistanz-Schwellenwert
- Zusammensetzung als Cluster-Adresse (oder Cluster-PLZ)

Probe	CT_50	CT_30	CT_10	PLZ
s1	1	1	1	1.1.1
s2	2	2	2	2.2.2
s3	2	2	3	2.2.3
s4	2	2	4	2.2.4
s5	2	2	4	2.2.4
s6	2	3	5	2.3.5
s7	2	3	5	2.3.5



Allel Qualitätskontrolle

- Gefundene Loci (bekannte + unbekannte Allele)
- Nicht-gefundene Loci (Gene werden durch Frameshift Mutationen nicht mehr erkannt)
- Entfernt Loci durch Filtern (z. B. Paraloge, zu lange oder zu kurze Allelsequenzen)
- Relativer Anteil gefundener Loci sollte **95%** betragen
- Mögliche Konsequenzen
 - Die Alleldistanzen zwischen einer Probe mit wenigen Loci und allen anderen Proben werden **unterschätzt**
 - Die Alleldistanzen zwischen anderen Proben bleibt davon unberührt
- Mögliche Ursachen
 - Probleme mit Qualität der Genomdaten
 - Genom ist unvollständig / zu stark fragmentiert
 - Genom passt nicht gut zum cgMLST Schema
- Details zu jeder Probe im chewieSnake Report



Report und Visualisierung mit grapetree

- Alle Ergebnisse der verschiedenen Bausteine von chewieSnake werden in einem interaktiven HTML report zusammengefasst
- Übersicht Tabs: Overview, Allele statistics, Allele distance table, Allele distance matrix, Clustering, Minimum-spanning tree, Links to files, Config and parameters, Help
- Beispiel Report auf gitlab: https://bfr_bioinformatics.gitlab.io/chewieSnake/report_chewiesnake.html
- Alternative: Visualisierung des Interaktiven mittels **grapetree**
 - Webseite https://achtman-lab.github.io/GrapeTree/MSTree_holder.html aufrufen
 - Vorberechneten Baum laden
 - Metadaten (tsv Datei) laden
 - Mehr Informationen unter <https://enterobase.readthedocs.io/en/latest/grapetree/grapetree-about.html>



Weitere Optionen

- --reads

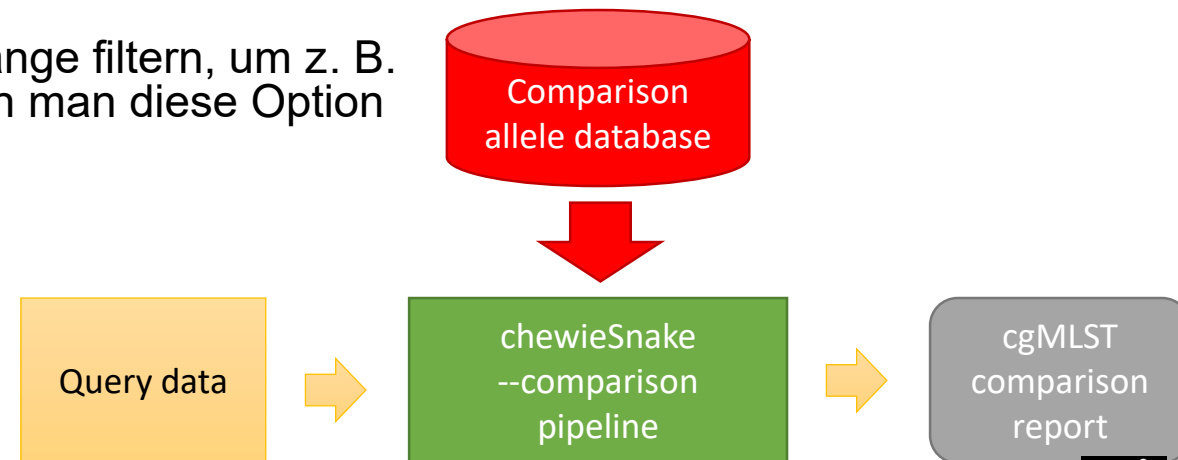
Falls keine Assembly vorhanden sind, kann chewieSnake Readdaten mit shovill zu Draftgenomen assemblieren. Bevorzugt ist aber das Vorschalten z. B. der AQUAMIS Pipeline, um die Sequenzierqualität besser bewerten zu können

- --frameshift_filter

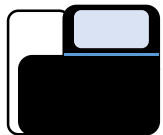
chewBBACA filtert nur neue Allele. Will man alle Allele nach Länge filtern, um z. B. Längendifferenzen nur bis zu 9 Basenpaaren zuzulassen, kann man diese Option benutzen.

- --comparison

Vergleich von neu berechneten (Query) Genomdaten zu (vorberechneten) Vergleichsdaten



Routineanwendung

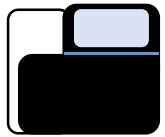


sample fq1

Sample1 fastq1_R1
Sample2 fastq2_R1
Sample3 fastq3_R1
Sample4 fastq4_R1



Routineanwendung



sample fq1

Sample1 fastq1_R1
Sample2 fastq2_R1
Sample3 fastq3_R1
Sample4 fastq4_R1



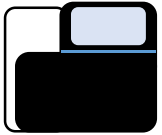
AQUAMIS
pipeline



AutoQC



Routineanwendung



sample fq1

Sample1 fastq1_R1
Sample2 fastq2_R1
Sample3 fastq3_R1
Sample4 fastq4_R1



AQUAMIS
pipeline



AutoQC

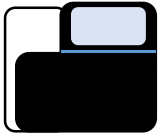
Assembly
sample
sheet

sample assembly

Sample1 fasta1
Sample2 fasta2
Sample3 fasta3
Sample4 fasta4



Routineanwendung



sample fq1

Sample1 fastq1_R1
Sample2 fastq2_R1
Sample3 fastq3_R1
Sample4 fastq4_R1



AQUAMIS
pipeline



AutoQC

Assembly
sample
sheet

sample assembly

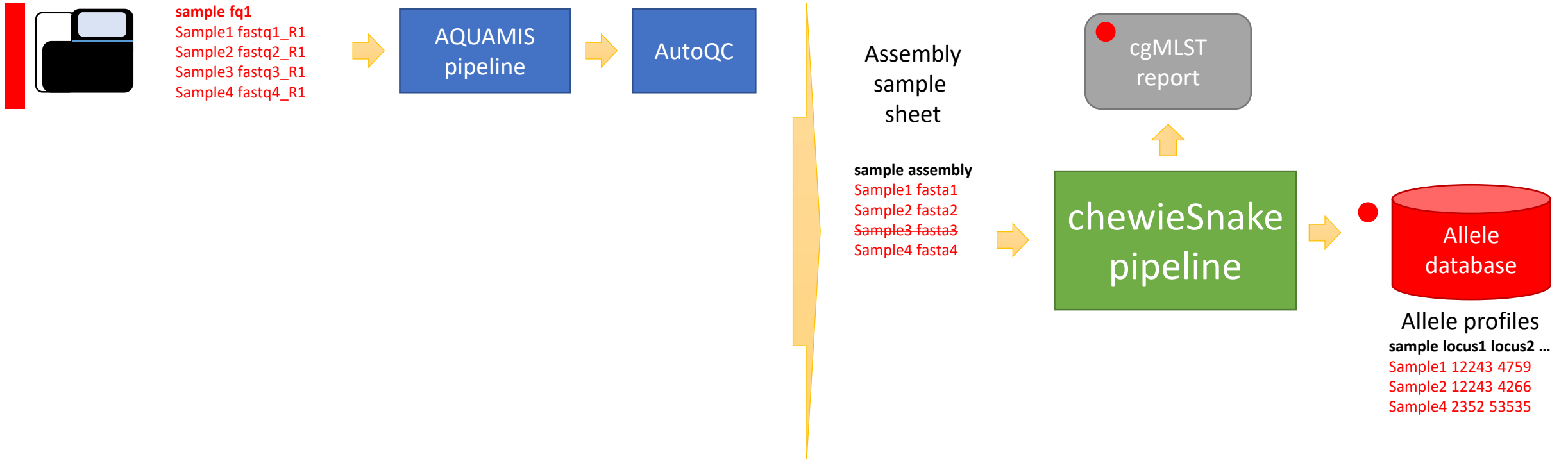
Sample1 fasta1
Sample2 fasta2
Sample3 fasta3
Sample4 fasta4



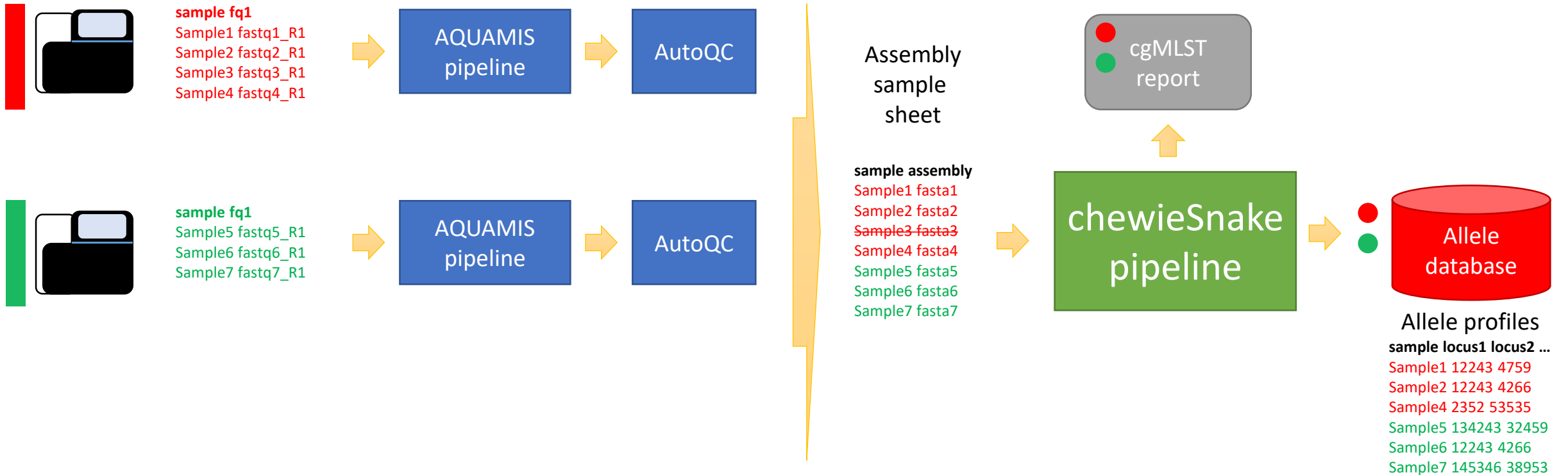
chewieSnake
pipeline



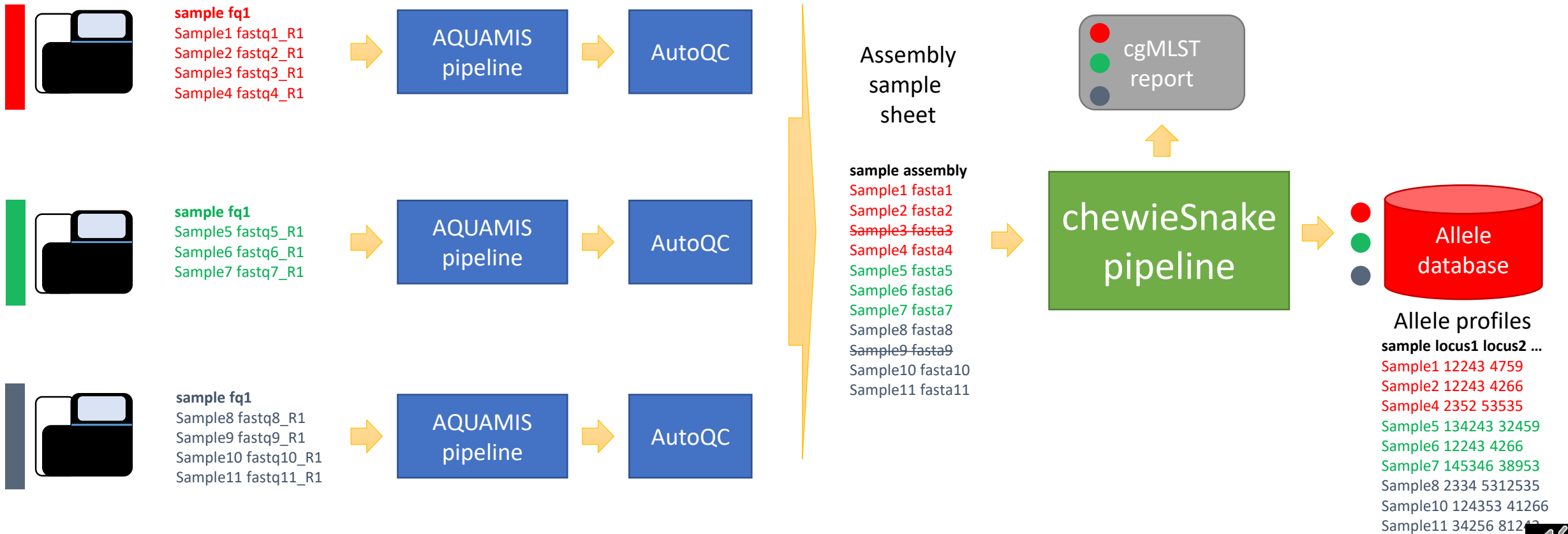
Routineanwendung



Routineanwendung



Routineanwendung



Zusammenfassung

- chewieSnake ist eine Allele calling cgMLST Pipeline
 - Probenmanagement und Aufruf der einzelnen Programmbausteine
 - Allele calling aller Proben mit chewBBACA
 - Zusammenführung aller Probandaten in Distanzmatrix, Minimum-spanning tree, Cluster PLZ
 - Darstellung der Ergebnisse in einem interaktiven HTML Report
- Mit Hilfe von chewieSnake kann eine Allele QC durchgeführt werden
- Die chewieSnake Ergebnisse sind mit anderen Laboren vergleichbar, da das Hashing Verfahren die gleichen Allelsequenzen in eindeutige und identische „allele hashes“ übersetzt



Weitere Informationen und Links

- Gitlab https://gitlab.com/bfr_bioinformatics/chewieSnake
 - Repository des freiverfügbaren und offener Quellcodes
 - Dokumentation der Software mit allen Programmoptionen
 - Beschreibung zur Installation
 - Quellen für öffentliche cgMLST Schemata
 - Tutorium: Beispielausführung anhand Testdatensatz
- Beispielreport: https://bfr_bioinformatics.gitlab.io/chewieSnake/report_chewiesnake.html
- Publikationen
 - *Salmonella* <https://doi.org/10.3389/fmicb.2021.649517>
 - *Listeria* <https://doi.org/10.1099/mgen.0.000491>
- Screencast 5: Anwendung von chewieSnake in GSS
- Screencast 8: Installation der Pipelines



GenoSalmSurv screencasts

das Projekt

1

die Pipelines

AQUAMIS und die QC Entscheidung

Allele calling mit chewieSnake

Gemeinsame GenoSalmSurv Datenbank

Metachewiereport + Interpretation

6

Demo einer Ausbruchsanalyse

7

Installationshilfe

8

