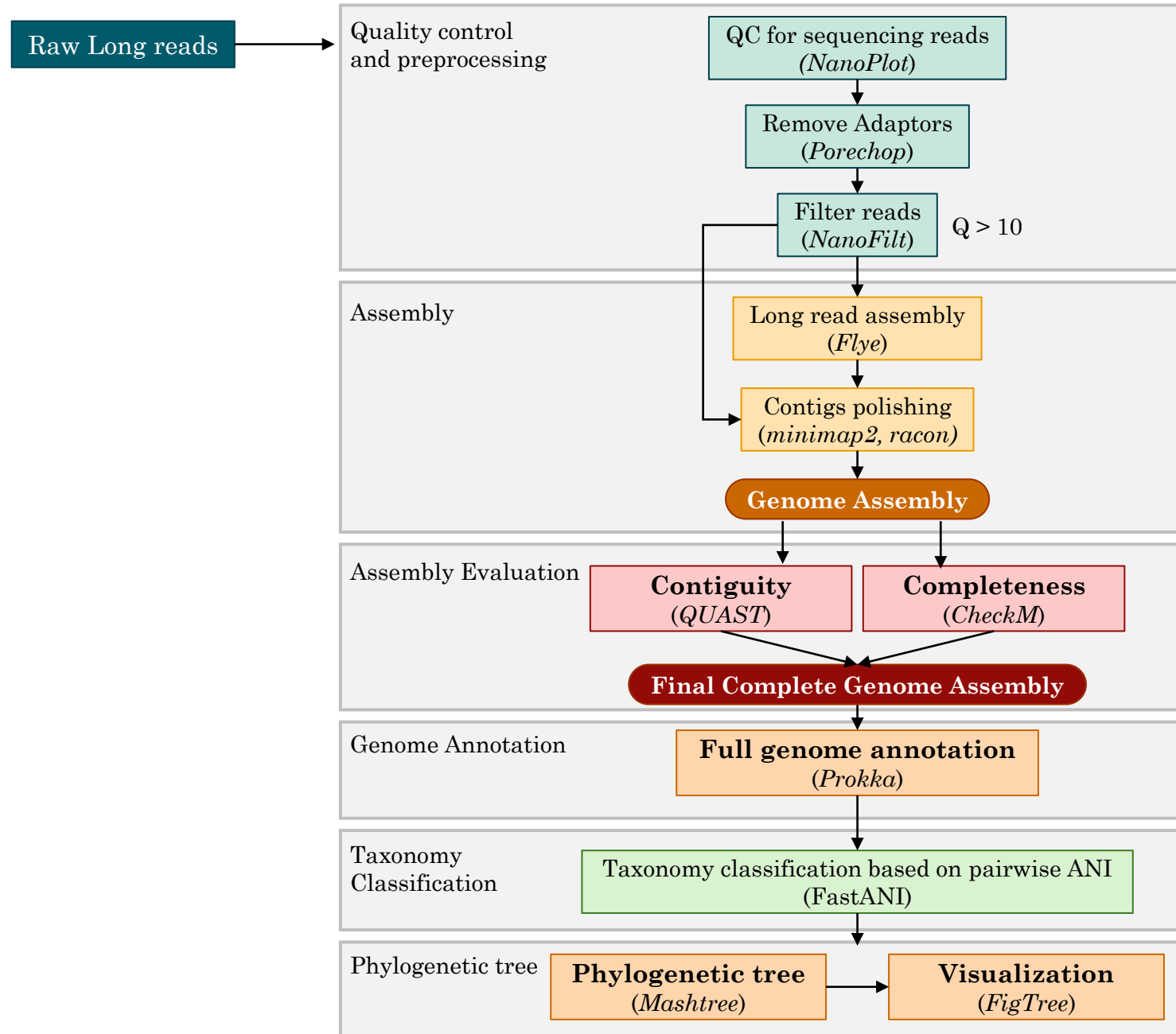




Session 3: Bacterial Genomic Analysis

Hands-on Bacterial Genome Analysis

Analysis Pipeline



Dataset Overview

Bartonella schoenbuchensis Strain L2, a bacterium that causes **bacteremia** in ruminants and is primarily transmitted by deer keds (*Lipoptena cervi*)

Bacteremia presence of bacteria in the blood, leading to infection and potentially serious health issues



Deer



Lipoptena cervi



Bartonella schoenbuchensis Strain L2

- Genome size: 1.669 Mb
- Level : Complete
- Release date: May, 2024
- Sequencing technology: ONT MinION
- GenBank accession: GCA_039555305.1
- SRA accession: SRR28636163



GENOME SEQUENCES
August 2024 Volume 13 Issue 8 e00397-24
<https://doi.org/10.1128/mra.00397-24>

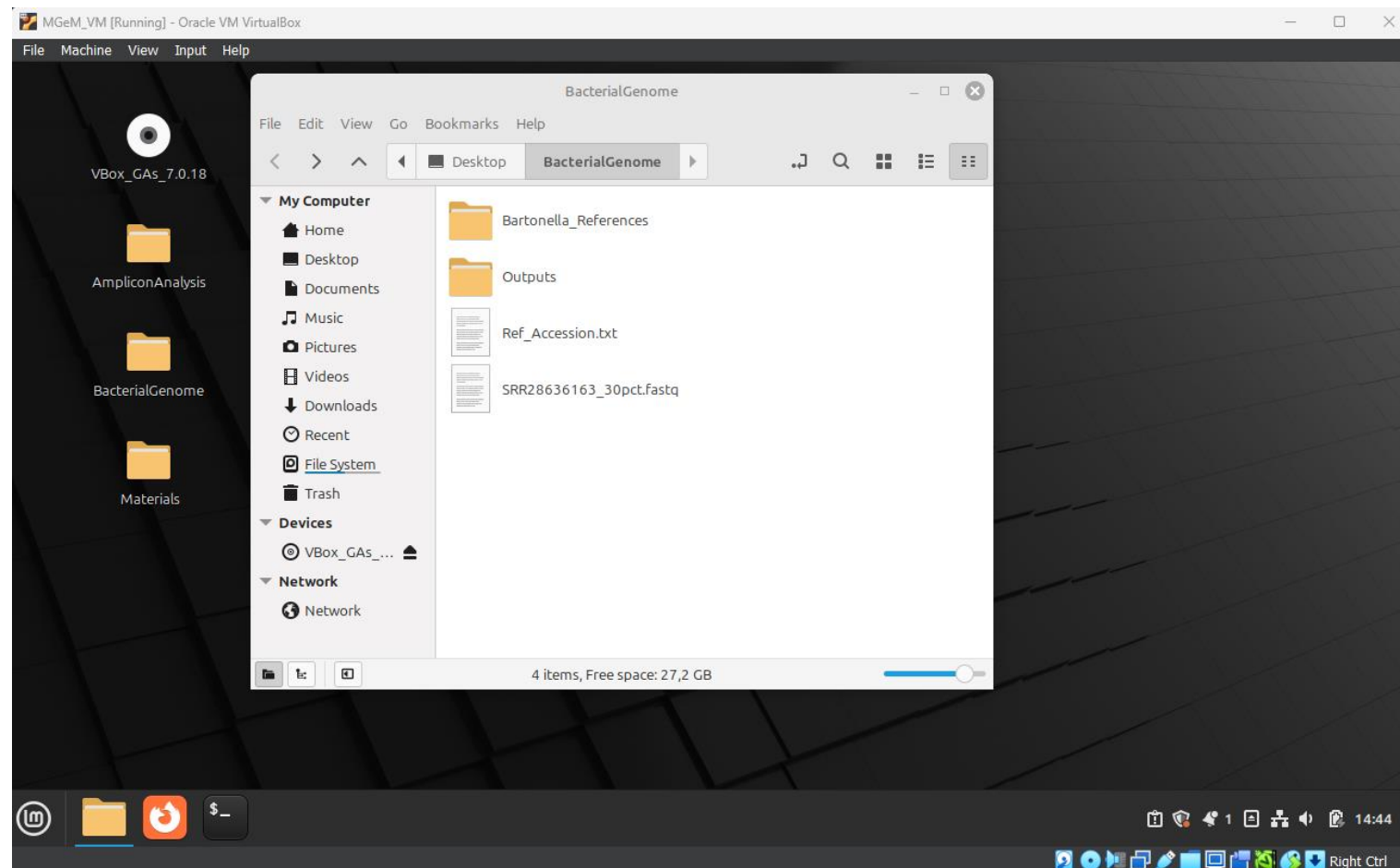
Genome sequences of three *Bartonella schoenbuchensis* strains from Czechia

Luis Solis Cayo^{1,2,3,4}, Iva Hammerbauerová⁵, Julian Sommer⁴, Zahra Nemati⁴, Wibke Ballhorn⁴, Pablo Tsukayama ², Alexander Dichter⁴, Jan Votýpka⁵, Volkhard A. J. Kempf ⁴

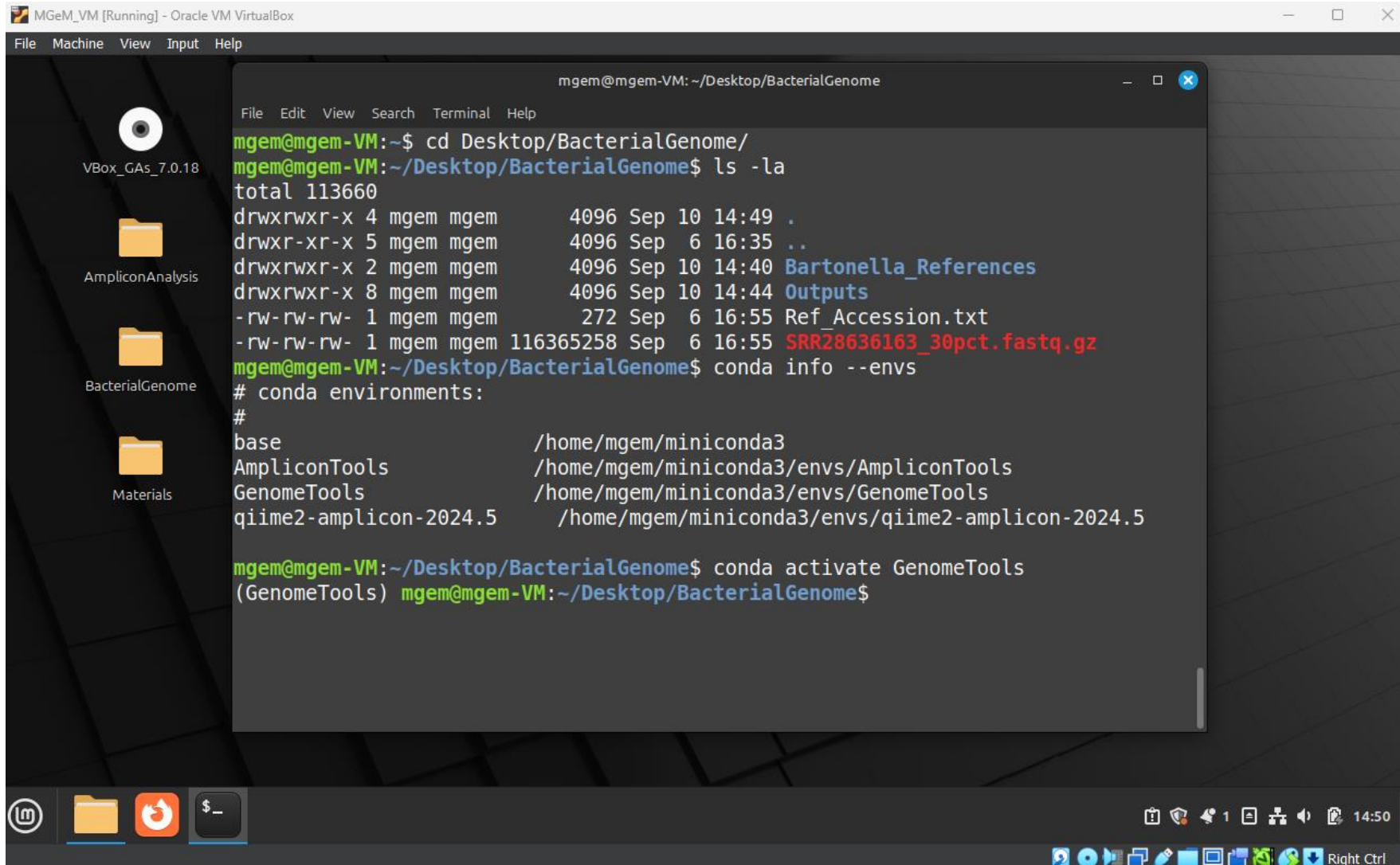
Starting point

Working directory:

`/home/mgem/Desktop/BacterialGenome/`



Starting point



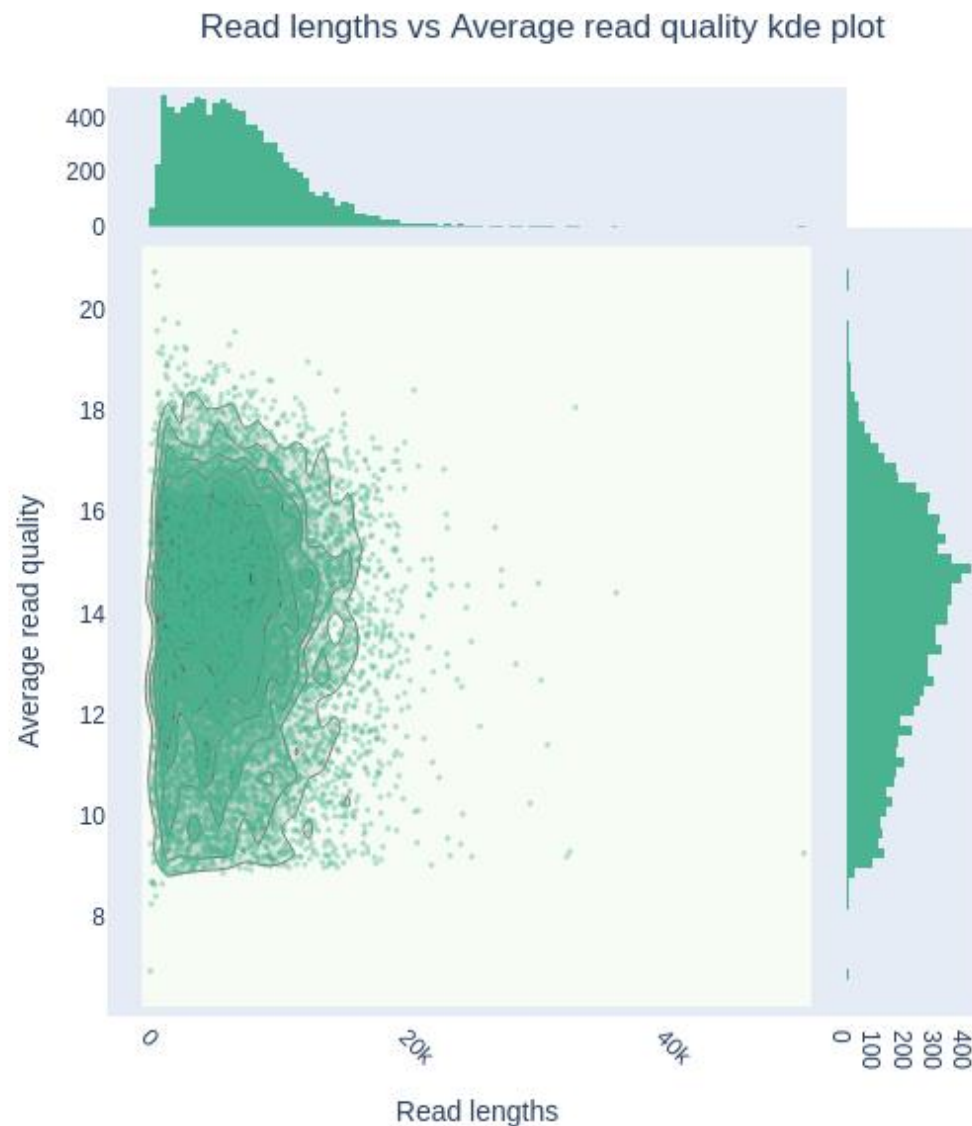
```
mgem@mgem-VM: ~/Desktop/BacterialGenome
File Edit View Search Terminal Help
mgem@mgem-VM:~$ cd Desktop/BacterialGenome/
mgem@mgem-VM:~/Desktop/BacterialGenome$ ls -la
total 113660
drwxrwxr-x 4 mgem mgem      4096 Sep 10 14:49 .
drwxr-xr-x 5 mgem mgem      4096 Sep  6 16:35 ..
drwxrwxr-x 2 mgem mgem      4096 Sep 10 14:40 Bartonella_References
drwxrwxr-x 8 mgem mgem      4096 Sep 10 14:44 Outputs
-rw-rw-rw- 1 mgem mgem       272 Sep  6 16:55 Ref_Accession.txt
-rw-rw-rw- 1 mgem mgem 116365258 Sep  6 16:55 SRR28636163_30pct.fastq.gz
mgem@mgem-VM:~/Desktop/BacterialGenome$ conda info --envs
# conda environments:
#
base                                /home/mgem/miniconda3
AmpliconTools                       /home/mgem/miniconda3/envs/AmpliconTools
GenomeTools                         /home/mgem/miniconda3/envs/GenomeTools
qiime2-amplicon-2024.5              /home/mgem/miniconda3/envs/qiime2-amplicon-2024.5

mgem@mgem-VM:~/Desktop/BacterialGenome$ conda activate GenomeTools
(GenomeTools) mgem@mgem-VM:~/Desktop/BacterialGenome$
```

Quality control

NanoPlot

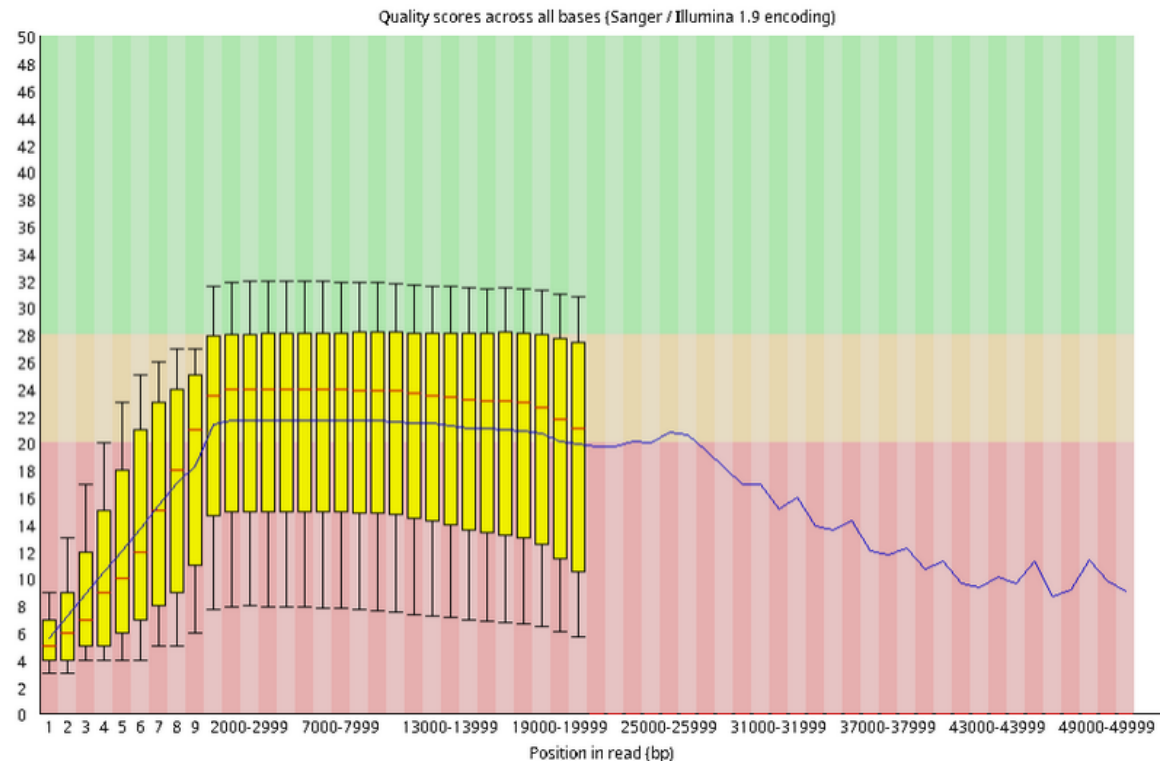
```
NanoStats.txt X
General summary:
Mean read length:      6,787.1
Mean read quality:     12.7
Median read length:    6,138.5
Median read quality:   14.0
Number of reads:       18,000.0
Read length N50:       8,992.0
STDEV read length:     4,303.6
Total bases:           122,168,691.0
Number, percentage and megabases of reads above quality cutoffs
>Q10:  16954 (94.2%) 115.3Mb
>Q15:   5790 (32.2%) 39.3Mb
>Q20:    4 (0.0%) 0.0Mb
>Q25:    0 (0.0%) 0.0Mb
>Q30:    0 (0.0%) 0.0Mb
Top 5 highest mean basecall quality scores and their read lengths
1:  20.7 (478)
2:  20.5 (696)
3:  20.3 (1209)
4:  20.1 (1281)
5:  19.9 (1141)
Top 5 longest reads and their mean basecall quality score
1: 50931 (9.3)
2: 36351 (14.4)
3: 36122 (10.0)
4: 33191 (18.1)
5: 32743 (9.3)
```



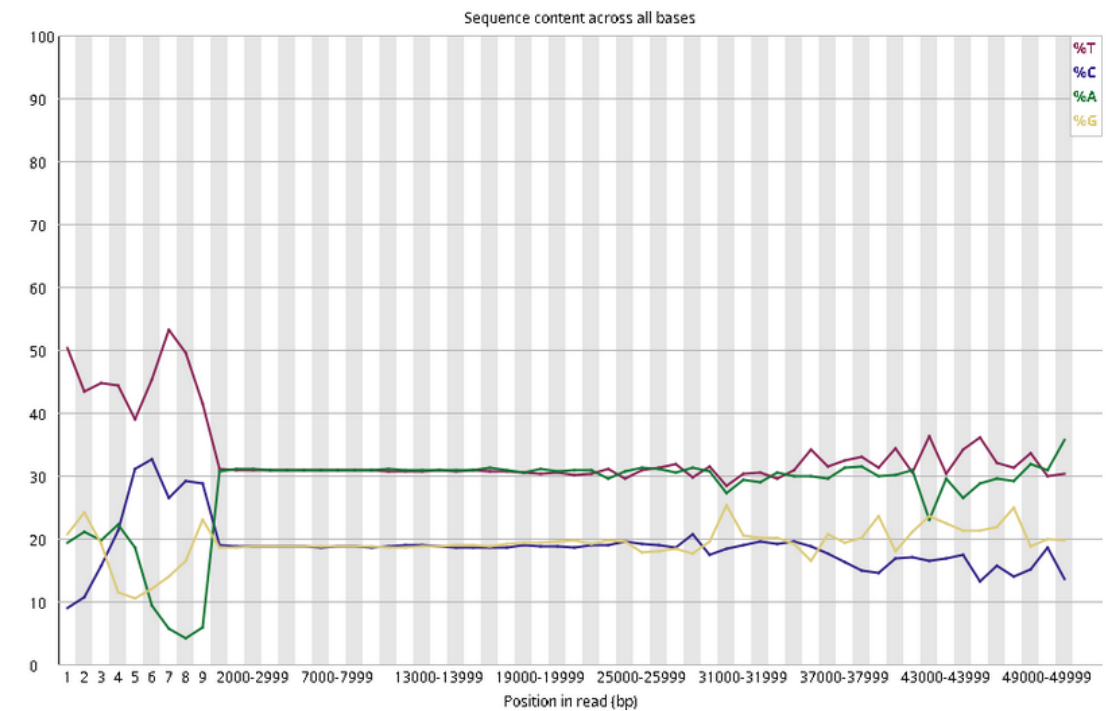
Quality control

FastQC

✖ Per base sequence quality



✖ Per base sequence content



Quality control: Adaptor trimming and filtering

Porechop

```
Trimming adapters from read ends
SQK-NSK007_Y_Top: AATGTACTTCGTTCAGTTACGTATTGCT
SQK-NSK007_Y_Bottom: GCAATACGTAACTGAACGAAGT
BC06_rev: GACTACTTTCTGCCTTTGCGAGAA
BC06: TTCTCGCAAAGGCAGAAAGTAGTC
BC06: TTCTCGCAAAGGCAGAAAGTAGTC
BC06_rev: GACTACTTTCTGCCTTTGCGAGAA
NB06_start: AATGTACTTCGTTCAGTTACGTATTGCTAAGGTTAAGACTACTTTCTGCCTTTGCGAGAACAGCACC
T
NB06_end: AGGTGCTGTTCTCGCAAAGGCAGAAAGTAGTCTTAACCTTAGCAATACGTAACTGAACGAAGT

18,000 / 18,000 (100.0%)

17,700 / 18,000 reads had adapters trimmed from their start (1,187,964 bp removed)
12,975 / 18,000 reads had adapters trimmed from their end (663,867 bp removed)
```

NanoFilt

Minimum Q-score = 10

Quality control: Adaptor trimming and filtering

Before

```
NanoStats.txt x
General summary:
Mean read length:      6,787.1
Mean read quality:     12.7
Median read length:    6,138.5
Median read quality:   14.0
Number of reads:       18,000.0
Read length N50:       8,992.0
STDEV read length:     4,303.6
Total bases:           122,168,691.0
Number, percentage and megabases of reads above quality cutoffs
>Q10:  16954 (94.2%) 115.3Mb
>Q15:   5790 (32.2%) 39.3Mb
>Q20:    4 (0.0%) 0.0Mb
>Q25:    0 (0.0%) 0.0Mb
>Q30:    0 (0.0%) 0.0Mb
Top 5 highest mean basecall quality scores and their read lengths
1:  20.7 (478)
2:  20.5 (696)
3:  20.3 (1209)
4:  20.1 (1281)
5:  19.9 (1141)
Top 5 longest reads and their mean basecall quality score
1: 50931 (9.3)
2: 36351 (14.4)
3: 36122 (10.0)
4: 33191 (18.1)
5: 32743 (9.3)
```



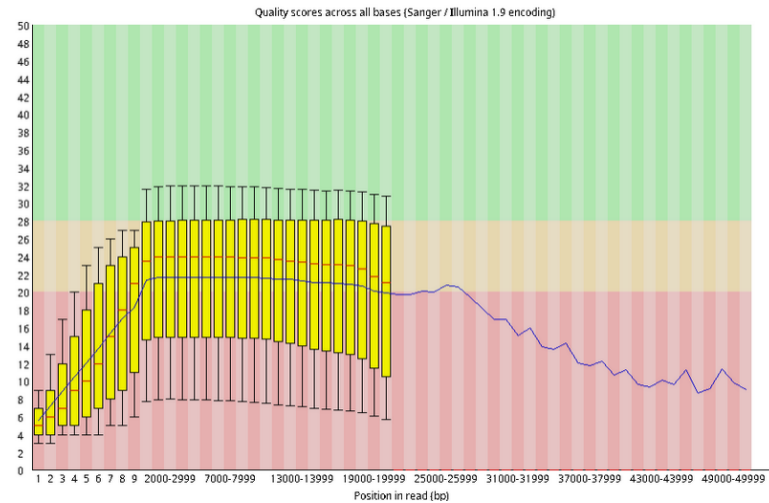
After

```
NanoStats.txt x
General summary:
Mean read length:      6,686.3
Mean read quality:     13.2
Median read length:    6,049.0
Median read quality:   14.2
Number of reads:       16,985.0
Read length N50:       8,903.0
STDEV read length:     4,254.5
Total bases:           113,567,424.0
Number, percentage and megabases of reads above quality cutoffs
>Q10:  16985 (100.0%) 113.6Mb
>Q15:   6105 (35.9%) 39.8Mb
>Q20:    15 (0.1%) 0.0Mb
>Q25:    0 (0.0%) 0.0Mb
>Q30:    0 (0.0%) 0.0Mb
Top 5 highest mean basecall quality scores and their read lengths
1:  22.7 (289)
2:  22.5 (132)
3:  22.2 (287)
4:  22.0 (566)
5:  21.8 (151)
Top 5 longest reads and their mean basecall quality score
1: 36231 (14.4)
2: 33124 (18.1)
3: 30935 (11.4)
4: 30578 (13.3)
5: 30234 (14.6)
```

Quality control: Adaptor trimming and filtering

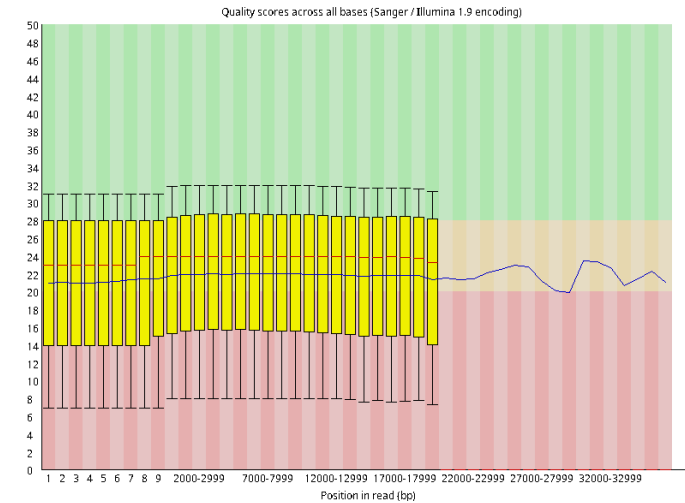
Before

✖ Per base sequence quality

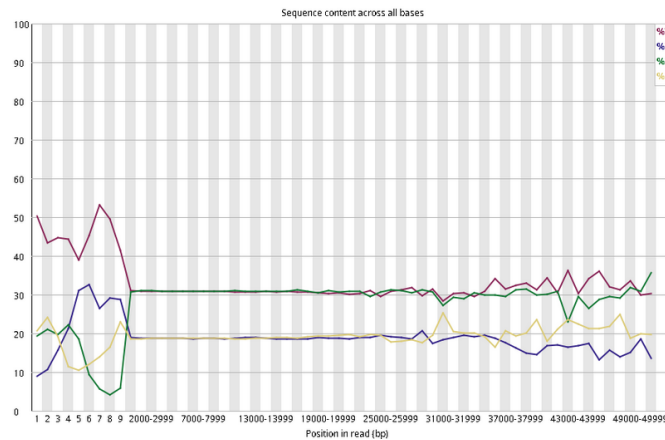


After

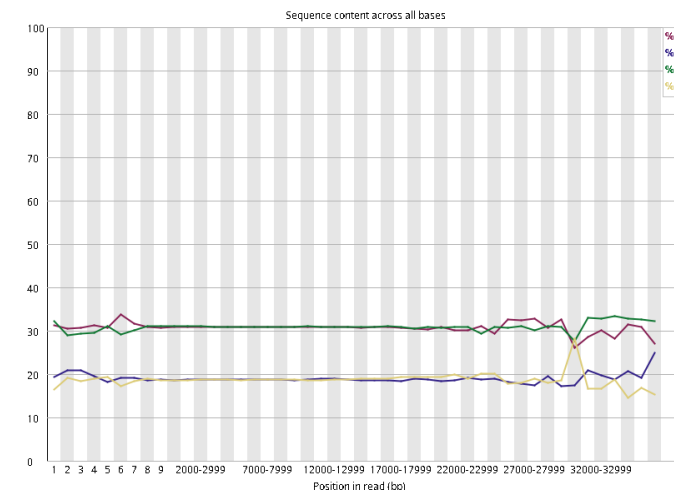
⚠ Per base sequence quality



✖ Per base sequence content



⚠ Per base sequence content



Assembly and Polishing

Flye

The assembly of the bacterial genome using the Flye assembler typically requires more than 2 hours to complete

To save time, **we will skip this step** during the session and proceed with the pre-assembled genome.



assembly.fasta ➡

```
assembly.fasta x
>contig_1
TACCGGTACGATCAGTGCTATGACGCTTTTTAAGGTAGGGGGCTTGGACGTTATACTGAG
CAATCTTTTGAGTGGTCAATATTCACCACATAATTTACTGGGAGCTGAGCAGATTTTCGTT
GGGAGGCGCTTCCAATGTTTCGTGGCAGCGTGAGAGTTTGCTTTTTGGCAATAACGGTTT
TTTTATCCGTAATGATTTGTCTTTGCGCACACTACCGTGGAATAACAACGCTTCTCTCAA
GAAGGTTTTAGGTGAACTGCGTCCTTATGTCGGTTTGGACTATGGGCGTGTTTTTCGCA
ACCCTTGTATGGGTTTACGCATGAACAGCTTGCCGGTTGGACGGCAGGGGTCAAGTTTGT
GGGGGGATAGTTTCCCTTGACGCAAGCTATTCCAGTGTTTTGTGGAGTACAGTTAAGCAC
AAGAAGCTTGGAACATTTTTTATGACATTAACAATGAATCTTTAAGTATATTATTGGAGG
ATACAAGCATGGGATATGAGCAGAGAGCAAAAAGAACAACATGGTTTGAAGTTTTAGTCT
```

flye.log

```
Total length: 1693842
Fragments: 3
Fragments N50: 1477396
Largest frg: 1477396
Scaffolds: 0
Mean coverage: 68
```

Assembly_info.txt

```
assembly_info.txt x
#seq_name length cov. circ. repeat mult. alt_group graph_path
contig_1 1477396 69 N N 1 * -4,-2,-2,-2,1,-4
contig_3 203114 66 N N 1 * -4,-2,-2,-2,-2,-2,3,-4
contig_4 13332 77 N Y 1 * 4
```

Assembly metrics and completeness

Quast

QUAST

Quality Assessment Tool for Genome Assemblies by [CAB](#)

10 September 2024, Tuesday, 09:08:54

[View in Icarus contig browser](#)

All statistics are based on contigs of size ≥ 500 bp, unless otherwise noted (e.g., "# contigs

Worst Median Best

☒ Show heatmap

Statistics without reference	assembly	assembly_polished
# contigs	3	3
# contigs (≥ 0 bp)	3	3
# contigs (≥ 1000 bp)	3	3
# contigs (≥ 5000 bp)	3	3
# contigs (≥ 10000 bp)	3	3
# contigs (≥ 25000 bp)	2	2
# contigs (≥ 50000 bp)	2	2
Largest contig	1 477 396	1 477 598
Total length	1 693 842	1 694 036
Total length (≥ 0 bp)	1 693 842	1 694 036
Total length (≥ 1000 bp)	1 693 842	1 694 036
Total length (≥ 5000 bp)	1 693 842	1 694 036
Total length (≥ 10000 bp)	1 693 842	1 694 036
Total length (≥ 25000 bp)	1 680 510	1 680 710
Total length (≥ 50000 bp)	1 680 510	1 680 710
N50	1 477 396	1 477 598
N90	203 114	203 112
auN	1 313 069	1 313 271
L50	1	1
L90	2	2
GC (%)	38.03	38.02

Assembly metrics and completeness

CheckM

Taxon list 

CheckM output (Terminal)



```
(GenomeTools) mgem@mgem-VM:~/Desktop/BacterialGenome$ checkm taxon_list
[2024-09-10 09:15:11] INFO: CheckM v1.2.3
[2024-09-10 09:15:11] INFO: checkm taxon_list
[2024-09-10 09:15:11] INFO: CheckM data: /home/mgem/miniconda3/envs/GenomeTools/checkm_data
[2024-09-10 09:15:11] INFO: [CheckM - taxon_list] Listing available taxonomic-specific marker sets.
```

Rank	Taxon	# genomes	# marker genes	# marker sets
life	Prokaryote	5656	56	24
domain	Archaea	207	149	107
domain	Bacteria	5449	104	58
phylum	Acidobacteria	15	399	276
phylum	Actinobacteria	731	204	119
phylum	Aquificae	18	486	369
phylum	Bacteroidetes	419	286	195
phylum	Chlamydiae	64	455	185
phylum	Chlorobi	12	612	333
phylum	Chloroflexi	20	225	149
phylum	Crenarchaeota	54	217	168
phylum	Cyanobacteria	129	472	368
phylum	Deferribacteres	6	564	339
phylum	Deinococcus-Thermus	40	528	359
phylum	Dictyoglomi	2	1060	103

Bin Id	Marker lineage	# genomes	# markers	# marker sets	0	1	2	3	4	5+	Completeness	Contamination	Strain heterogeneity
assembly_polished	Bartonella (5)	29	708	188	19	688	1	0	0	0	97.46	0.02	0.00
assembly	Bartonella (5)	29	708	188	29	677	2	0	0	0	95.36	0.55	0.00

Genome Annotation

Prokka

Annotation

File Edit View Go Bookmarks Help

< > ^

BacterialGenome Annotation

My Computer

Home

Desktop

Documents

Music

Pictures

Videos

Downloads

Recent

File System

Trash

Name	Size	Type
L2.err	452,9 kB	Text
L2.faa	538,8 kB	Text
L2.ffn	1,5 MB	Text
L2.fna	1,7 MB	Text
L2.fsa	1,7 MB	Text
L2.gbkg	3,5 MB	Text
L2.gff	2,1 MB	Text
L2.log	31,6 kB	Text
L2.sqn	5,8 MB	Text
L2.tbl	331,6 kB	Text
L2.tsv	111,0 kB	Text
L2.txt	94 bytes	Text

```
##gff-version 3
##sequence-region contig_1 1 1477598
##sequence-region contig_3 1 203112
##sequence-region contig_4 1 13326
contig_1 Prodigal:002006 CDS 20 379 . + 0 ID=EGBDFEDO_00001;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00001;product=
contig_1 Prodigal:002006 CDS 488 5347 . + 0 ID=EGBDFEDO_00002;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00002;product=
contig_1 Prodigal:002006 CDS 5404 7251 . + 0 ID=EGBDFEDO_00003;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00003;product=
contig_1 Prodigal:002006 CDS 7389 8534 . + 0 ID=EGBDFEDO_00004;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00004;product=
contig_1 Prodigal:002006 CDS 8531 8881 . + 0 ID=EGBDFEDO_00005;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00005;product=
contig_1 Prodigal:002006 CDS 8919 10055 . + 0 ID=EGBDFEDO_00006;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00006;product=
contig_1 Prodigal:002006 CDS 10052 10453 . + 0 ID=EGBDFEDO_00007;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00007;product=
contig_1 Prodigal:002006 CDS 10551 11636 . + 0 ID=EGBDFEDO_00008;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00008;product=
contig_1 Prodigal:002006 CDS 11662 11985 . + 0 ID=EGBDFEDO_00009;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00009;product=
contig_1 Prodigal:002006 CDS 12269 12409 . - 0 ID=EGBDFEDO_00010;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00010;product=
contig_1 Prodigal:002006 CDS 12617 12805 . - 0 ID=EGBDFEDO_00011;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00011;product=
contig_1 Prodigal:002006 CDS 12747 12992 . - 0 ID=EGBDFEDO_00012;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00012;product=
contig_1 Prodigal:002006 CDS 13019 13138 . + 0 ID=EGBDFEDO_00013;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00013;product=
contig_1 Aragorn:001002 tRNA 13452 13527 . - ID=EGBDFEDO_00014;inference=COORDINATES:profile:Aragorn:001002;locus_tag=EGBDFEDO_00014;product=
contig_1 Prodigal:002006 CDS 14002 14805 . + 0 ID=EGBDFEDO_00015;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00015;product=
sequence:UniProtKB:P0AEC0;locus_tag=EGBDFEDO_00015;note=UPF053 inner membrane protein YoaE;product=hypothetical protein
contig_1 Prodigal:002006 CDS 14874 15572 . + 0 ID=EGBDFEDO_00016;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00016;product=
sequence:UniProtKB:P0AEC0;locus_tag=EGBDFEDO_00016;note=UPF053 inner membrane protein YoaE;product=hypothetical protein
contig_1 Prodigal:002006 CDS 15948 16892 . - 0 ID=EGBDFEDO_00017;Name=ybhN;db_xref=COG:COG0392;gene=ybhN;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00017;product=
sequence:UniProtKB:P75770;locus_tag=EGBDFEDO_00017;product=Inner membrane protein YbhN
contig_1 Prodigal:002006 CDS 18520 20001 . - 0 ID=EGBDFEDO_00018;Name=secDF;db_xref=COG:COG0341;gene=secDF;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00018;product=
sequence:UniProtKB:Q5SKE6;locus_tag=EGBDFEDO_00018;product=Protein translocase subunit SecDF
contig_1 Prodigal:002006 CDS 19986 21020 . - 0 ID=EGBDFEDO_00019;Name=secD;db_xref=COG:COG0342;gene=secD;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00019;product=
sequence:UniProtKB:P0AG90;locus_tag=EGBDFEDO_00019;product=Protein translocase subunit SecD
```

	A	B	C	D	E	F	G
1	locus_tag	ftype	length_bp	gene	EC_number	COG	product
116	EGBDFEDO_00115	CDS	966	argS_1	6.1.1.19	COG0018	Arginine--tRNA ligase
117	EGBDFEDO_00116	CDS	849	argS_2	6.1.1.19		Arginine--tRNA ligase
118	EGBDFEDO_00117	CDS	1200				Deoxyguanosinetriphosphate triphosphohydrolase-like protein
119	EGBDFEDO_00118	CDS	324	sufA_1		COG0316	Protein SufA
120	EGBDFEDO_00119	CDS	1005				hypothetical protein
121	EGBDFEDO_00120	CDS	789				hypothetical protein
122	EGBDFEDO_00121	CDS	1350	glmU			Bifunctional protein GlmU
123	EGBDFEDO_00122	CDS	1824	nodM	2.6.1.16		Glutamine--fructose-6-phosphate aminotransferase [isomerizing]
124	EGBDFEDO_00123	CDS	1908	recG	3.6.4.12	COG1200	ATP-dependent DNA helicase RecG
125	EGBDFEDO_00124	CDS	513	dnaE1_1	2.7.7.7		DNA polymerase III subunit alpha
126	EGBDFEDO_00125	CDS	2673	dnaE1_2	2.7.7.7		DNA polymerase III subunit alpha
127	EGBDFEDO_00126	CDS	333				hypothetical protein
128	EGBDFEDO_00127	CDS	1248	mhbM	1.14.13.24	COG0654	3-hydroxybenzoate 6-hydroxylase
129	EGBDFEDO_00128	CDS	249				hypothetical protein
130	EGBDFEDO_00129	CDS	2139	ppk	2.7.4.1		Polyphosphate kinase
131	EGBDFEDO_00130	CDS	216				hypothetical protein
132	EGBDFEDO_00131	CDS	1314	gppA	3.6.1.40		Guanosine-5'-triphosphate-3'-diphosphate pyrophosphatase
133	EGBDFEDO_00132	CDS	825	rdm_1	3.1.13.5	COG0349	Ribonuclease D
134	EGBDFEDO_00133	CDS	1791	aspS	6.1.1.12		Aspartate--tRNA ligase
135	EGBDFEDO_00134	CDS	2250	parC	5.6.2.2	COG0188	DNA topoisomerase 4 subunit A
136	EGBDFEDO_00135	CDS	1248	bcr			Bicyclomycin resistance protein
137	EGBDFEDO_00136	CDS	333	grxD		COG0278	Glutaredoxin 4

L2.txt x

organism: Genus species strain
contigs: 3
bases: 1694036
CDS: 1749
rRNA: 6
tRNA: 42
tmRNA: 1

Genome Annotation

General Feature Format (GFF)

Contig ID	Source	Type	Start	End	Score	Strand	Phase	Attributes
contig_1	Prodigal:002006	CDS	20	379	.	+	0	ID=EGBDFEDO_00001;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00001;product=hypothetical protein
contig_1	Prodigal:002006	CDS	488	5347	.	+	0	ID=EGBDFEDO_00002;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00002;product=hypothetical protein
contig_1	Prodigal:002006	CDS	5404	7251	.	+	0	ID=EGBDFEDO_00003;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00003;product=hypothetical protein
contig_1	Prodigal:002006	CDS	7389	8534	.	+	0	ID=EGBDFEDO_00004;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00004;product=hypothetical protein
contig_1	Prodigal:002006	CDS	8531	8881	.	+	0	ID=EGBDFEDO_00005;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00005;product=hypothetical protein
contig_1	Prodigal:002006	CDS	8919	10055	.	+	0	ID=EGBDFEDO_00006;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00006;product=hypothetical protein
contig_1	Prodigal:002006	CDS	10052	10453	.	+	0	ID=EGBDFEDO_00007;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00007;product=hypothetical protein
contig_1	Prodigal:002006	CDS	10551	11636	.	+	0	ID=EGBDFEDO_00008;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00008;product=hypothetical protein
contig_1	Prodigal:002006	CDS	11662	11985	.	+	0	ID=EGBDFEDO_00009;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00009;product=hypothetical protein
contig_1	Prodigal:002006	CDS	12269	12409	.	-	0	ID=EGBDFEDO_00010;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00010;product=hypothetical protein
contig_1	Prodigal:002006	CDS	12617	12805	.	-	0	ID=EGBDFEDO_00011;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00011;product=hypothetical protein
contig_1	Prodigal:002006	CDS	12747	12992	.	-	0	ID=EGBDFEDO_00012;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00012;product=hypothetical protein
contig_1	Prodigal:002006	CDS	13019	13138	.	+	0	ID=EGBDFEDO_00013;inference=ab initio prediction:Prodigal:002006;locus_tag=EGBDFEDO_00013;product=hypothetical protein
contig_1	Aragorn:001002	tRNA	13452	13527	.	-	.	ID=EGBDFEDO_00014;inference=COORDINATES:profile:Aragorn:001002;locus_tag=EGBDFEDO_00014;product=tRNA-Lys(ctt)

- **ContigID:** The name of the sequence where the feature is located
- **Source:** The algorithm or procedure that generated the feature
- **Type:** The feature type name, like “CDS” or “tRNA”
- **Start:** Genomic start position of the feature
- **End:** Genomic end position of the feature
- **Score:** Confidence of the source in the annotated feature
- **Strand:** A single character indicates the strand of the feature. This can be “+” (positive, or 5’->3’), “-“, (negative, or 3’->5’), “.” (undetermined)
- **Phase:** Phase of CDS features; it can be either one of 0, 1, 2 (for CDS features) or “.” (for everything else)
- **Attributes:** A list of tag-value pairs separated by a semicolon with additional information about the feature.

Genome Annotation

Prokka

Annotation

File Edit View Go Bookmarks Help

< > ^ ◀ ▶ BacterialGenome Annotation

My Computer

Home

Desktop

Documents

Music

Pictures

Videos

Downloads

Recent

File System

Trash

Name	Size	Type
L2.err	452,9 kB	Text
L2.faa	538,8 kB	Text
L2.ffn	1,5 MB	Text
L2.fna	1,7 MB	Text
L2.fsa	1,7 MB	Text
L2.gbk	3,5 MB	Text
L2.gff	2,1 MB	Text
L2.log	31,6 kB	Text
L2.sqn	5,8 MB	Text
L2.tbl	331,6 kB	Text
L2.tsv	111,0 kB	Text
L2.txt	94 bytes	Text

L2.gbk x

LOCUS contig_1 1477598 bp DNA linear 10-SEP-2024

DEFINITION Genus species strain strain.

ACCESSION

VERSION

KEYWORDS .

SOURCE Genus species

ORGANISM Genus species

Unclassified.

COMMENT Annotated using prokka 1.14.6 from <https://github.com/tseemann/prokka>.

FEATURES

Location/Qualifiers

source 1..1477598

/organism="Genus species"

/mol_type="genomic DNA"

/strain="strain"

CDS 20..379

/locus_tag="EGBDFED0_00001"

/inference="ab initio prediction:Prodigal:002006"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/translation="MTLFKVGGLDVILSNLLSGQYSPHNLLGAEQISLGGASNVRGTR

ESLLFGNNGFFIRNDLSRLTLPWNNNASLKKVLGELRPYVGLDYGRVFSQPLYGF

THE QLAGWTAGVKFVGDSFP"

...

ORIGIN

1 taccggtacg atcagtgcata tgacgctttt taaggtaggg ggcttggacg ttatactgag

61 caatcttttg agtgggtcaat attcaccaca taatttactg ggagctgagc agatttcggt

121 gggaggcgct tccaatgttc gtggcacgcg tgagagtgtt ctttttgga ataacgggtt

181 ttttatccgt aatgatttgt ctttgcgcac actaccgtgg aataacaacg cttctctcaa

241 gaaggtttta ggtgaactgc gtccttatgt cggtttgac tatgggcgtg ttttttcgca

301 acccttgat gggtttacgc atgaacagct tgccggttgg acggcagggg tcaagtttgt

361 gggggatagt ttcccttgac gcaagctatt ccagtgttt gtggagtaca gttaagcaca

421 agaagcttgg aacatttttt atgacattaa caatgaatct ttaagtatat tattggagga

Header starting with the LOCUS line

Feature listing any annotated features like CDS

The actual nucleotide sequence, ending with a // line

GenBank format (gbk)

Genome Annotation

Prokka

Annotation

File Edit View Go Bookmarks Help

< > ^

BacterialGenome

Annotation

My Computer

Home

Desktop

Documents

Music

Pictures

Videos

Downloads

Recent

File System

Trash

Name	Size	Type
L2.err	452,8 kB	Text
L2.faa	538,8 kB	Text
L2.ffn	1,5 MB	Text
L2.fna	1,7 MB	Text
L2.fsa	1,7 MB	Text
L2.gbk	3,5 MB	Text
L2.gff	2,1 MB	Text
L2.log	31,6 kB	Text
L2.sqn	5,8 MB	Text
L2.tbl	331,6 kB	Text
L2.tsv	111,0 kB	Text
L2.txt	94 bytes	Text

L2.faa x L2.ffn x

```
>EGBDFED0_00001 hypothetical protein
MTLFKVGGLDVILSNLLSGQYSPHNLLGAEQISLGGASNVRGTRESLLFGNNGFFIRNDL
SLRTL PWNNNASLKKVLGELRPYVGLDYGRVFSQPLYGFTHEQLAGWTAGVKFVGDSFP
>EGBDFED0_00002 hypothetical protein
MGYEQRAKRTTWFEVLVSSTILKKIFLGGIGLPFLLQPATLQAQIAVDPNAGAAHRPDIV
AAPNGVPSIDIVTPNGKGLSHNKYYNFIIGNPGVILNNHAQEVGQSQLGGIMPGNPHLRY
SGSAKVILNEVTSGKRSALHGPAEVFGRQADVIIANPNGISCDGCGFINTPHATLTTGVP
EIDGSGFLKGFVKGDDITFGPQGANFFSGKGAVDIVDIVSRTVHFEGSVAGREIGVTAG
TGHFDYASRQMKELTDITGKPEYAIDGSVLGALQADKIKLVATEKGVGVRMRHDMANAG
QLHLSADGKISLKDIFGHDGVVLQSKSDTLHAKHVTSKKHVEVAACKGVTLEAVGADGHL
```

L2.faa x L2.ffn x

```
>EGBDFED0_00001 hypothetical protein
ATGACGCTTTTAAAGGTAGGGGGCTTGGACGTTATACTGAGCAATCTTTGAGTGGTCAA
TATTCACCACATAATTTACTGGGAGCTGAGCAGATTTCTGTTGGGAGGCGCTTCCAATGTT
CGTGGCACGCGTGAGAGTTTGCTTTTGGCAATAACGGTTTTTTTATCCGTAATGATTTG
TCTTTGCGCACACTACCGTGAATAACAACGCTTCTCTCAAGAAGTTTTAGGTGAACG
CGTCCTTATGTCGGTTTGGACTATGGGCGTGTTTTTCGCAACCCTTGTATGGGTTTACG
CATGAACAGCTTGCCGTTGGACGGCAGGGGTCAAGTTTGTGGGGGATAGTTTCCCTTGA
>EGBDFED0_00002 hypothetical protein
ATGGGATATGAGCAGAGAGCAAAAAGAACAACATGGTTTGAAGTTTTAGTCTCCAGTACG
ATACTCAAAAAGATTTTTCTAGGCGGGATTGGTTTACCTTTTCTTTTACAACCAGCAACG
CTACAGGCGCAAATTGCTGTTGATCCTAACGCAGGTGCTGCTCATCGTCCTGATATTGTT
GCAGCGCCCAATGGAGTTCCTCGATTGATATTGTTACACCAAATGGCAAAGGCTTATCA
CACAATAAATATTACAATTTTAATATTGGCAATCCAGGCGTCATTCTCAACAATCATGCG
CAAGAAGTGGGACAATCGCAATTGGGAGGCATTATGCCGGGCAATCCGCATTTGCGCTAT
TCTGGTTTCGGCAAAAAGTGATTTTAAATGAAGTGACCAGTGGCAAGCGCAGTGCTCTTCAC
GGTCCAGCGGAGGTTTTTGGTCGCCAGGCAGATGTGATTATAGCCAATCCCAACGGCATA
```


Taxonomy Classification

FastANI

Path to reference genomes →

```
GenomeList.txt ×
/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_000015445.1_ASM1544v1_genomic.fna
/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_000022725.1_ASM2272v1_genomic.fna
/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_000196435.1_ASM19643v1_genomic.fna
/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_000253015.1_ASM25301v1_genomic.fna
/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_000341355.1_ASM34135v1_genomic.fna
/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_001281405.1_ASM128140v1_genomic.fna
/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_003606325.2_ASM360632v3_genomic.fna
/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_003606345.3_ASM360634v3_genomic.fna
/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_009936175.1_ASM993617v1_genomic.fna
/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_013388295.1_ASM1338829v1_genomic.fna
/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_019930925.1_ASM1993092v1_genomic.fna
/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_022559585.1_ASM2255958v1_genomic.fna
/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_023920085.1_ASM2392008v1_genomic.fna
/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_024297065.1_ASM2429706v1_genomic.fna
/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_039555305.1_BS-L2_genomic.fna
/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_900638615.1_58174_F01_genomic.fna
/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_900638635.1_58174_G01_genomic.fna
```

Metadata of reference genomes (metadata.tsv)



	A	B	C	D	E	F	G	H	I
1	assembly_accession	bioproject	biosample	refseq_category	relation_to_type_material	taxid	species_taxid	organism_name	infraspecific_name
2	GCF_009936175.1	PRJNA224116	SAMD00165657	representative genome	na	803	803	Bartonella quintana	strain=MF1-1
3	GCF_900638615.1	PRJNA224116	SAMEA104307704	representative genome	na	807	807	Bartonella elizabethae	strain=NCTC12898
4	GCF_023920085.1	PRJNA224116	SAMN21428279	representative genome	na	33046	33046	Bartonella taylorii	strain=IBS296
5	GCF_900638635.1	PRJNA224116	SAMEA104307705	representative genome	na	33047	33047	Bartonella vinsonii	strain=NCTC12905
6	GCF_019930925.1	PRJNA224116	SAMN20888879	representative genome	assembly from type material	38323	38323	Bartonella henselae	strain=FDAARGOS_1462
7	GCF_013388295.1	PRJNA224116	SAMN15350025	representative genome	assembly from type material	52764	52764	Bartonella alsatica	strain=CIP 105477
8	GCF_039555305.1	PRJNA224116	SAMN40935016	representative genome	na	165694	165694	Bartonella schoenbuchensis	strain=L2
9	GCF_000015445.1	PRJNA224116	SAMN02604047	representative genome	assembly from type material	360095	774	Bartonella bacilliformis KC583	strain=KC583; ATCC 35685
10	GCF_000196435.1	PRJNA224116	SAMEA3138339	representative genome	assembly from type material	382640	85701	Bartonella tribocorum CIP 105476	strain=type strain:506 = CIP 105476
11	GCF_000022725.1	PRJNA224116	SAMN02603206	representative genome	na	634504	33045	Bartonella grahamii as4aup	strain=as4aup
12	GCF_000253015.1	PRJNA224116	SAMEA2272739	representative genome	na	696125	56426	Bartonella clarridgeiae 73	strain=73
13	GCF_000341355.1	PRJNA224116	SAMN02604270	representative genome	assembly from type material	1094489	388640	Bartonella australis AUST/NH1	strain=AUST/NH1
14	GCF_001281405.1	PRJNA224116	SAMN02803869	representative genome	assembly from type material	1318743	1318743	Bartonella ancashensis	strain=20.00
15	GCF_003606325.2	PRJNA224116	SAMN08765113	representative genome	assembly from type material	2133959	2133959	Bartonella kosoyi	strain=Tel Aviy
16	GCF_003606345.3	PRJNA224116	SAMN08765112	representative genome	assembly from type material	2267275	2267275	Bartonella krasnovii	strain=OE 1-1
17	GCF_022559585.1	PRJNA224116	SAMN23010494	representative genome	na	2893471	2893471	Bartonella machadoae	strain=46A
18	GCF_024297065.1	PRJNA224116	SAMN29470399	representative genome	assembly from type material	2961895	2961895	Bartonella harrusi	strain=117A

Taxonomy Classification

FastANI

FastaANI output (fastani_out.tsv)



ANI



1	assembly_polished.fasta	/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_039555305.1_BS-L2_genomic.fna	99.9596	552	563
2	assembly_polished.fasta	/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_000253015.1_ASM25301v1_genomic.fna	80.691	341	563
3	assembly_polished.fasta	/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_000015445.1_ASM1544v1_genomic.fna	80.4129	348	563
4	assembly_polished.fasta	/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_023920085.1_ASM2392008v1_genomic.fna	80.2882	363	563
5	assembly_polished.fasta	/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_009936175.1_ASM993617v1_genomic.fna	80.1594	334	563
6	assembly_polished.fasta	/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_019930925.1_ASM1993092v1_genomic.fna	80.0279	344	563
7	assembly_polished.fasta	/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_013388295.1_ASM1338829v1_genomic.fna	79.9861	331	563
8	assembly_polished.fasta	/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_900638635.1_58174_G01_genomic.fna	79.9436	331	563
9	assembly_polished.fasta	/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_000022725.1_ASM2272v1_genomic.fna	79.4769	357	563
10	assembly_polished.fasta	/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_022559585.1_ASM2255958v1_genomic.fna	79.4428	335	563
11	assembly_polished.fasta	/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_003606325.2_ASM360632v3_genomic.fna	79.4069	316	563
12	assembly_polished.fasta	/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_024297065.1_ASM2429706v1_genomic.fna	79.3653	335	563
13	assembly_polished.fasta	/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_900638615.1_58174_F01_genomic.fna	79.3185	299	563
14	assembly_polished.fasta	/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_000196435.1_ASM19643v1_genomic.fna	79.1909	322	563
15	assembly_polished.fasta	/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_000341355.1_ASM34135v1_genomic.fna	79.1607	266	563
16	assembly_polished.fasta	/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_003606345.3_ASM360634v3_genomic.fna	79.1427	313	563
17	assembly_polished.fasta	/home/mgem/Desktop/BacterialGenome/Bartonella_References/GCF_001281405.1_ASM128140v1_genomic.fna	78.3844	229	563

Phylogenetic tree

Figtree

