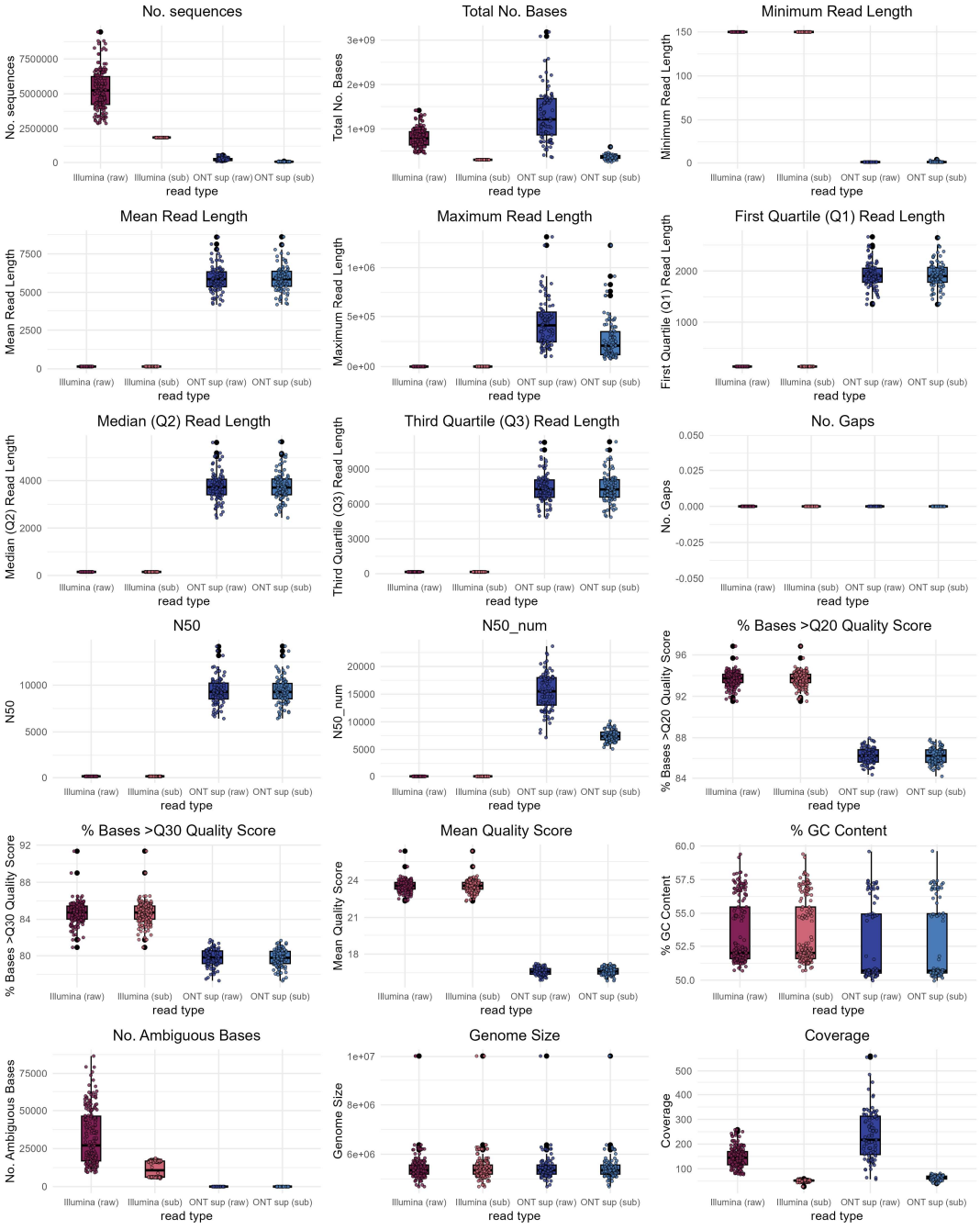
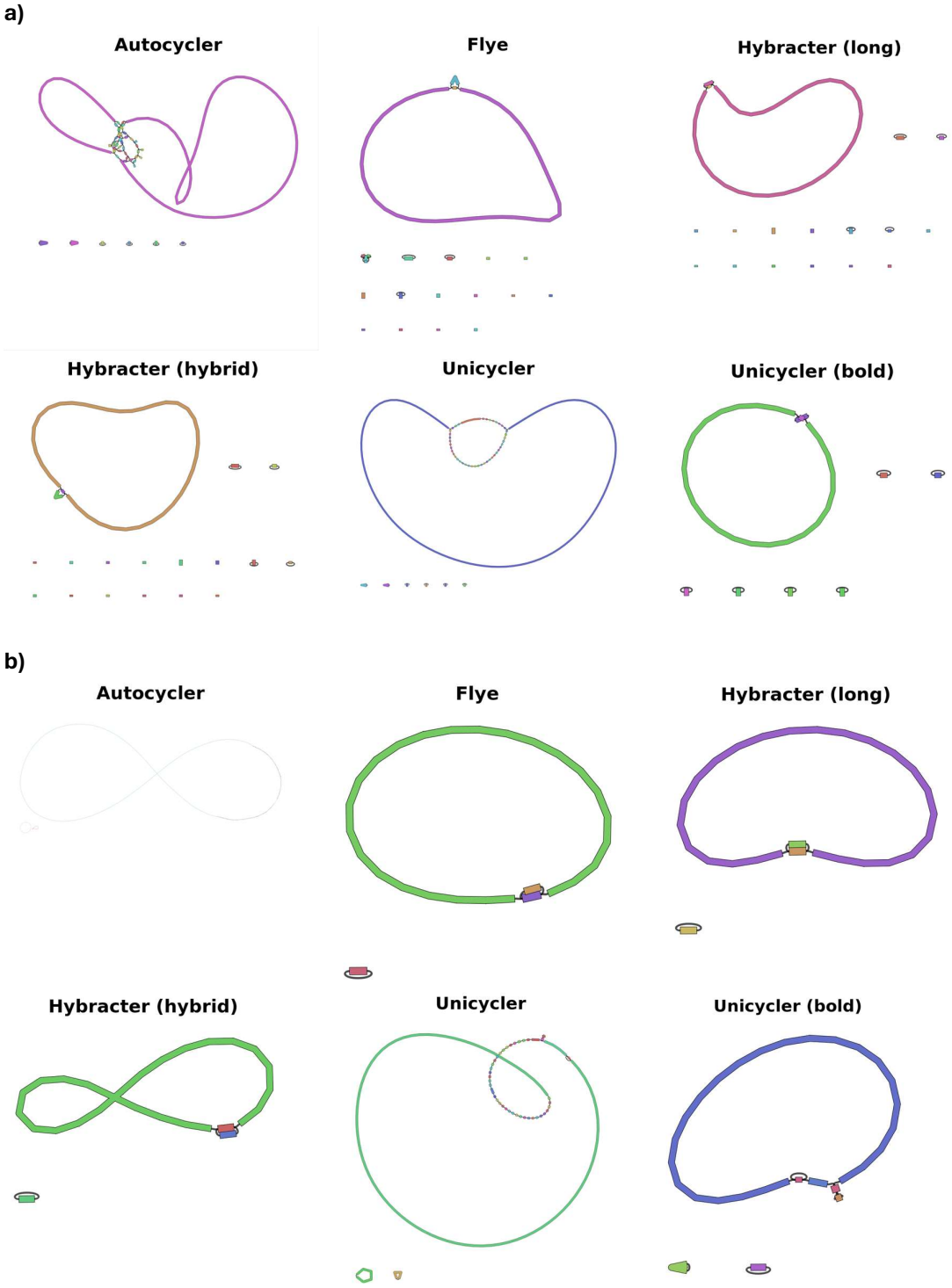
Supplementary Figures

Supplementary Figure S1: Quality control metrics of raw and subsampled Illumina short-reads and Dorado v5.0.0 super accurate basecalled Nanopore long-reads. Showing long-read subsampled set 1 (of 4) for the 92 pure culture isolates. N50 and N50_num (or L50) are both measures of sequence contiguity(1). N50 is the sequence length of the shortest contig at 50% of the total assembly length. N50_num is defined as the count of the smallest number of contigs whose added length makes up at least half of genome size.

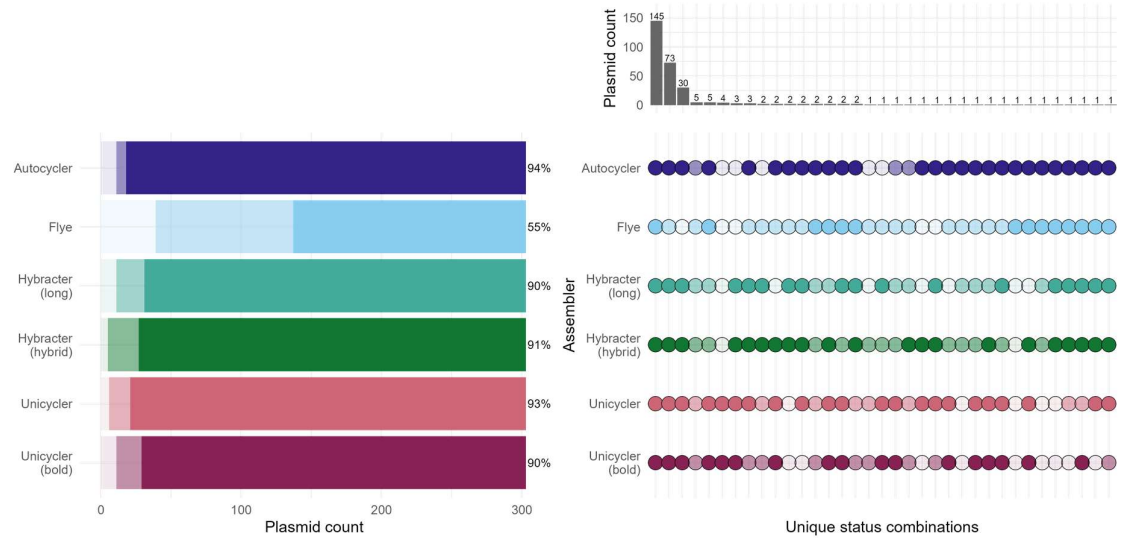


Supplementary Figure S2: Bandage plots of the two isolates where the chromosome was incompletely circularised by all assemblers. In both cases **a)** and **b),** there is a small unresolved repeat in Flye and Hybracter assemblies, which are more fragmented in Unicycler (normal; bold) assemblies, and very highly fragmented in Autocycler consensus assemblies. Note the Autocycler plot for panel b is so highly fragmented that the line representing the chromosome is very thin and barely visible.

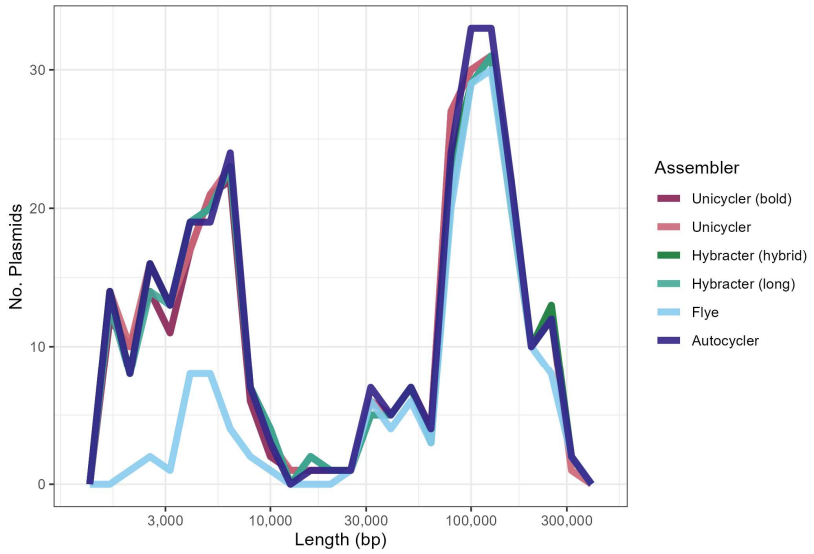


Supplementary Figure S3: Plasmid sequence reconstruction for 92 Enterobacterales isolates by different long-read only and hybrid assemblers, using the manually-curated consensus 'reference' plasmid set (n=303 plasmids). Reference plasmids in the manually curated set are circular contigs between 1,000-400,000bp in length that are present in at least 2 assemblers with a matching length (±10%) and mash distance (<0.025). **a)** UpSet plot showing assembly status combinations of plasmids across assemblers. Dark circles/bars indicate 'present' plasmids where length (±10%), mash distance (<0.025) and circularity all matched the 'reference' plasmid, lighter colours indicate misassembled plasmids, where the length difference was >10%, mash distance >0.025, or the contig was non-circular and the palest shades indicate absent plasmids, where no contig was found matching other plasmids in the reference plasmid set. **b)** Frequency polygon of length distribution of 'present' plasmids by assembler.

a)

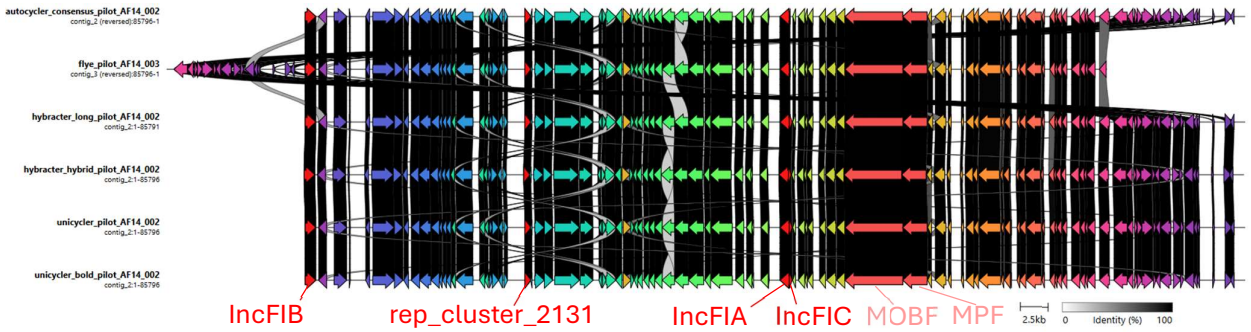


b)

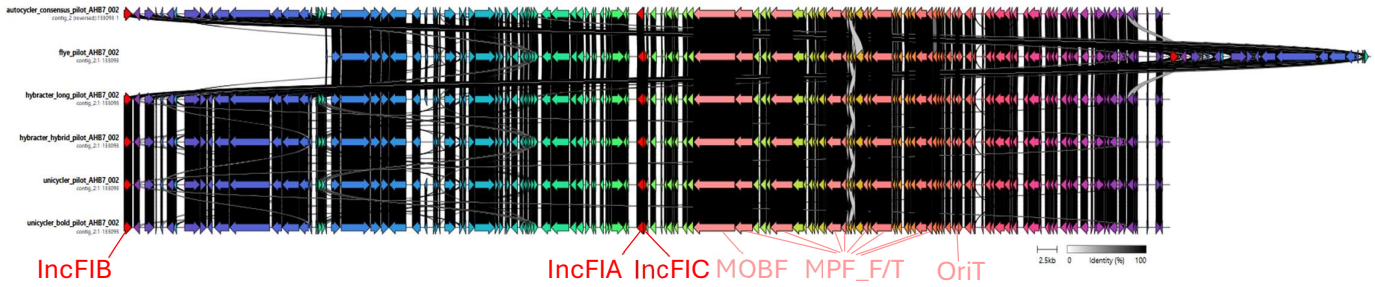


Supplementary Figure S4: Clinker plots of highly similar plasmids with different MOB-suite annotations. Replicon annotations are shown in bright red and labelled. Other mobility- and replication-associated plasmid machinery are shown in pale red and labelled. **a)** An 85,796bp IncFIA, IncFIB, IncFIC, rep_cluster_2131 plasmid sequence (isolate AF14) with a missing IncFIC annotation in the Autocycler and Flye assemblies (top 2), despite a mash distance of 0 between Autocycler and Hybracter (hybrid) assemblies. **b)** A 133,309bp IncFIA, IncFIB, IncFIC plasmid sequence (isolate AHB7) with the IncFIC replicon annotation missing from the Autocycler plasmid sequence, despite a mash distance of 0 between the Autocycler and Hybracter (hybrid) plasmid sequences. Note the Autocycler plasmid sequence is reversed and the Flye plasmid has a different starting point for both plasmids. The Flye plasmid is also reversed in a) compared to the bottom 4 assemblers' plasmids.

a)



b)



Supplementary Table S1: Species of the 92 pure culture Enterobacterales isolates, as assigned by Kraken2(2).

Species	Count (percentage)
Escherichia coli	58 (63%)
Klebsiella pneumoniae	21 (23%)
Klebsiella oxytoca	6 (7%)
Klebsiella aerogenes	2 (2%)
Enterobacter hormaechei	2 (2%)
Citrobacter freundii	1 (1%)
Citrobacter portucalensis	1 (1%)
Serratia marcescens	1 (1%)

Supplementary Table S2: Raw and subsampled sequencing read metrics for Illumina short-read and Nanopore long-read sequences for 92 pure culture Enterobacterales isolates.

	Raw reads	Subsampled reads
	Median (IQR)	
Read depth (x genome)		
Short-read	290 (232-340)	104 (100-108)
Long-read	217 (158-313)	64 (59-70)
Read length		
Short-read	150 (150-150)	150 (150-150)
Long-read	5858 (5366-6338)	5849 (5398-6370)
Read quality (Q score)		
Short-read	23.6 (23.3-23.8)	23.6 (23.3-23.8)
Long-read	16.6 (16.4-16.8)	16.6 (16.4-16.8)

Supplementary Table S3: Plasmid reconstruction accuracy of different long-read only and hybrid assemblers for Dorado v5.0.0 super accurate basecalled Nanopore long-reads.
Plasmid reconstruction is compared to a manually-curated reference set of 'consensus' plasmids (n=303), where 'consensus' plasmids were circular contigs 1,000-400,000bp in length present across at least 2 assemblers with a similar length (±10%) and close mash distance (<0.025).

	Assembler						p-value†
	Autocycler n (%)	Flye n (%)	Hybracter (long) n (%)	Hybracter (hybrid) n (%)	Unicycler n (%)	Unicycler (bold) n (%)	
Present* plasmids	285 (94.1%)	166 (54.8%)	272 (89.8%)	276 (91.1%)	282 (93.1%)	274 (90.4%)	<0.0001
Misassembled plasmids**							
Non-circular	0 (0%)	18 (5.9%)	12 (4.0%)	13 (4.3%)	5 (1.7%)	3 (1%)	
Length mismatch	7 (2.3%)	50 (16.5%)	1 (0.3%)	2 (0.7%)	6 (2.0%)	7 (2.3%)	
Non-circular and length mismatch	0 (0%)	30 (9.9%)	7 (2.3%)	7 (2.3%)	4 (1.3%)	8 (2.6%)	
Absent plasmids	11 (3.6%)	39 (12.9%)	11 (3.6%)	5 (1.7%)	6 (2.0%)	11 (3.6%)	

*'Present' plasmids are defined as contigs 1,000-400,000bp in length meeting all three match criteria: circular, length (±10%) and mash distance (<0.025) of a the manually curated reference set of plasmids.

**Misassembled plasmids are defined as contigs that failed to meet at least 1 of the matching criteria, but could still be matched to the reference set based on a more distant mash distance.

***Absent plasmids were cases where only the circularity matched, or where, for an assembler, no contig could be matched to the rest of the reference plasmids match set based on mash distance.

†p-value for Fleiss' Kappa test for uneven proportions of 'present' plasmids across all assemblers.

Supplementary Table S4: Nucleotide-level accuracy of 12 assembler-polisher combinations (7 long-read only, 5 hybrid). Read-alignment metrics were derived by aligning Illumina short-reads to each assembler-polisher combination and variant calling with Feebayes(3) from Pypolca. Mean gene length is derived from CheckM2(4) output files. 7-locus MLST is annotated by mlst(5), and key resistance, virulence and stress genes by AMRFinder Plus(6).

		Autocycler				Flye				Hybracter		Unicycler		p-value*
		None	Medaka (subsampled)	Medaka (un-subsampled)	Polypolish +Pypolca	None	Medaka (subsampled)	Medaka (un-subsampled)	Polypolish +Pypolca	Long	Hybrid	Normal	Bold	
MLST														**<0.0001**
MLST	**(N=91)**	90 (99%)††	90 (99%)††	90 (99%)††	90 (99%)††	90 (99%)††	91 (100%)	91 (100%)	91 (100%)	91 (100%)	91 (100%)	91 (100%)	91 (100%)	
Read-alignment metrics														
SNV /Mb	**Median (IQR)**	0 (0-0.17)	0 (0-0)	0 (0-0)	0 (0-0)	0.18 (0-1.17)	0 (0-0.52)	0 (0-0.7)	0 (0-0)	0.2 (0-1.26)	0 (0-0)	0 (0-0.37)	0 (0-0.37)	**<0.0001**
	Range	0-6.54	0-7.45	0-5.27	0-3.09	0-10.81	0-35.38	0-41.79	0-4.08	0-35.37	0-10.41	0-4	0-4	
Indels /Mb	**Median (IQR)**	0.18 (0-0.39)	0 (0-0.2)	0 (0-0.19)	0 (0-0)	0.57 (0.19-1.13)	0.18 (0-0.51)	0.17 (0-0.36)	0 (0-0)	0.39 (0.19-0.75)	0 (0-0)	0 (0-0.2)	0 (0-0.34)	**<0.0001**
	Range	0-9.5	0-17.11	0-13.12	0-5.45	0-34.11	0-18.66	0-22.35	0-4.47	0-16.71	0-12.42	0-16.21	0-16.21	
QV	**Median (IQR)**	67 (63-100)	100 (64-100)	100 (64-100)	100 (100-100)	61 (57-67)	67 (60-100)	67 (61-100)	100 (100-100)	60 (58-64)	100 (100-100)	67 (62-100)	67 (62-100)	**<0.0001**
	Range	48.8-100	47.3-100	48.4-100	50.7-100	43.48-100	42.7-100	41.9-100	51.7-100	42.8-100	46.4-100	46.9-100	46.9-100	
CheckM2														
Mean Gene Length	**Median (IQR)**	312 (309-316)	312 (309-316)	312 (309-316)	312 (309-316)	312 (309-316)	312 (308-315)	312 (308-316)	312 (309-315)	312 (309-316)	312 (309-316)	312 (309-316)	312 (309-316)	**<0.0001**
	Range	300-323	300-323	300-323	300-323	299-323	299-323	299-323	299-323	300-323	300-323	298-324	300-324	
AMR Finder Plus														
AMR	**Median (IQR)**	4 (1-7)	4 (1-7)	4 (1-7)	4(1-7)	4 (1-7)	4 (1-7)	4 (1-7)	4 (1-7)	4 (1-7)	4 (1-7)	3 (1-7)	4 (1-7)	0.209
	Range	0-18	0-18	0-18	0-18	0-18	0-18	0-18	0-18	0-17	0-18	0-17	0-17	
Stress	**Median (IQR)**	1 (0-3)	1 (0-3)	1 (0-3)	1 (0-3)	1 (0-3)	1 (0-3)	1 (0-3)	1 (0-3)	1 (0-3)	1 (0-3)	1 (0-2)	1 (0-2)	0.687
	Range	0-26	0-26	0-26	0-26	0-26	0-26	0-26	0-26	0-26	0-26	0-26	0-26	
Virulence	**Median (IQR)**	1 (0-7)	1 (0-7)	1 (0-7)	1 (0-7)	1 (0-6)	1 (0-6)	1 (0-6)	1 (0-6)	1 (0-6)	1 (0-6)	1 (0-6)	1 (0-6)	0.736
	Range	0-35	0-35	0-35	0-35	0-35	0-35	0-35	0-35	0-35	0-35	0-35	0-35	

*p-value for Fleiss' Kappa test for uneven proportions of isolates with correct MLST profiles annotated across all assemblers, or Friedman's test for global differences in continuous variables across all assemblers.

†MLST typing schemes were only available for 91/92 pure culture isolates. The excluded sample was identified as *Serratia marcescens*.

†† The incorrectly assigned MLST in one isolate by autocycler consensus assemblies, with or without polishing, was due to duplication of one of the seven housekeeping genes (gyrB(10,10)).

References

1. Wikipedia. N50, L50, and related statistics 2024 [Available from: https://en.wikipedia.org/wiki/N50,_L50,_and_related_statistics].
2. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biology. 2019;20(1):257.10.1186/s13059-019-1891-0
3. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv. 2012:e1207.3907
4. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Nature Methods. 2023;20(8):1203-12.10.1038/s41592-023-01940-w
5. Seemann, Torsten. mlst. 2.23.0 ed: Github.
6. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci Rep. 2021;11(1):12728.10.1038/s41598-021-91456-0