A Supplementary Data: Arredondo-Alonso et al. On the (im)possibility to reconstruct plasmids from whole genome short-read sequencing data

A.1 Workflow

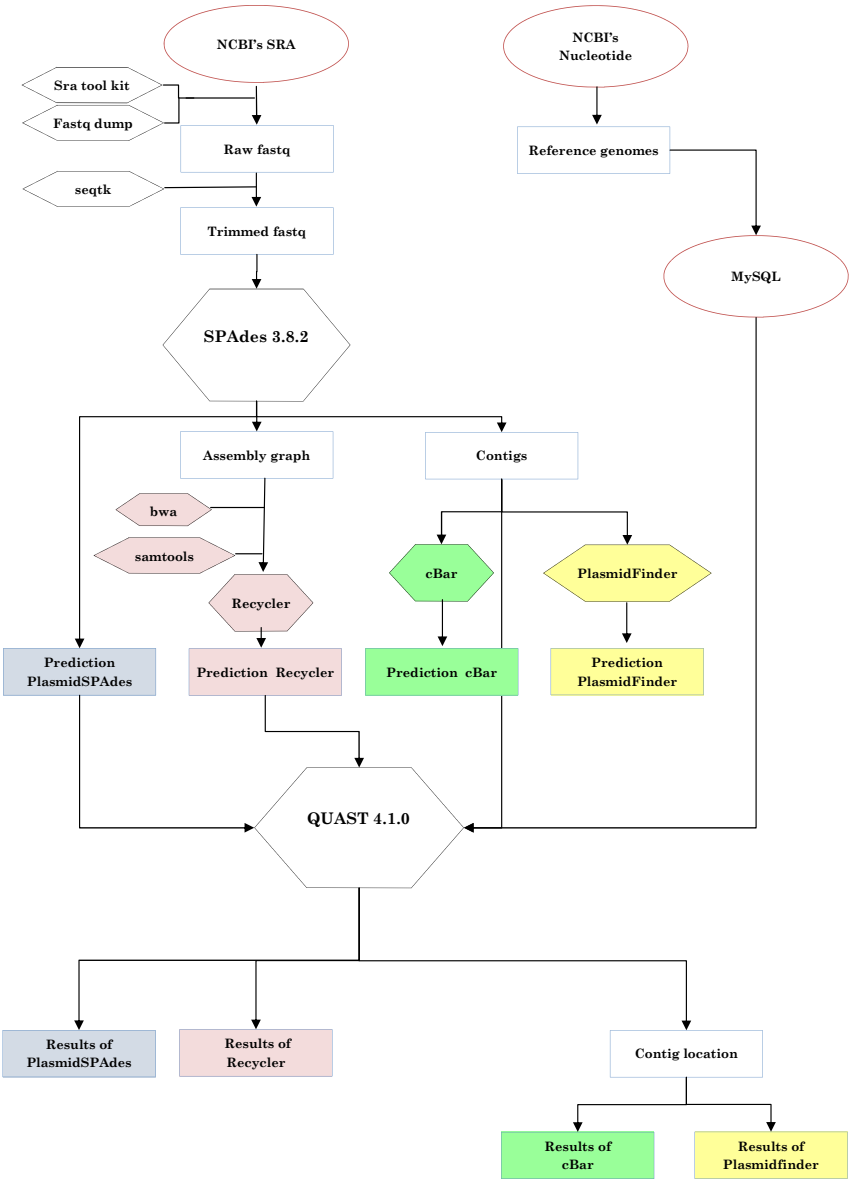

Figure 1. Process followed to analyze each of the 42 genomes. The result of each program was further analyzed using custom python and R scripts, available at git@gitlab.com:sirarredondo/Plasmid_Assembly.git.

A.2 Detailed description of genomes considered as positive controls

The following genomes were previously analyzed by the authors of PlasmidSPAdes and Recycler to validate their algorithms [16, 17].

A.2.1 Escherichia coli JJ1886

Recycler predicted the sequence of seven possible plasmids from *E.coli* JJ1886. Four of them corresponded to the reference plasmids whereas three sequences did not map either the plasmid references or the chromosome of *E.coli* JJ1886 (Figure 2). These three sequences were confirmed as plasmids by nucleotide BLAST although no other evidence of plasmid-related genes were found in the annotation by Prokka. In addition, we found two sequences of 42.6 kbp and 8.2 kbp corresponding to the chromosome of *E.coli* JJ1886. Annotation of the two sequences mapping to the chromosome suggested a phage origin of the 42.6 kbp sequence.

PlasmidSPAdes was able to recover a fraction of the plasmid pJJ1886_5, although plasmids pJJ1886_1 and pJJ1886_3 were not detected (Figure 2). Different components did not map to either the chromosome or the reference plasmids and were considered as putative unidentified plasmids. The components corresponding to *S.aureus* plasmids were present in a small copy number. Frequently, short-read length plasmids are in high copy number to ensure their prevalence in the next generations [27]. These findings suggested the sequences unidentified in *E.coli* JJ1886 corresponded to contamination during the library preparation. PlasmidSPAdes did not remove some parts of the chromosome from *S. aureus* because its coverage differed from *E.coli* JJ1886.

cBar identified several plasmid sequences as chromosomal resulting in a low precision (Table 1). However, it was the program with the best recall value with 83.27 because it recovered 12 contigs (>500 bp) belonging to pJJ1886_5. PlasmidFinder detected the presence of two plasmid replicon initiator sequences corresponding to the incompatibility group IncF. Both replicons are located in the plasmid pJJ1886_5 of the contigs with a size of 8.0 kbp and 12.9 kbp (Figure 2).

Table 1. Precision and recall of each program in *E. coli* JJ1886

Program	**Precision (%)**	**Recall (%)**
pSPAdes	56.49	45.95
Recycler	57.33	38.00
PlasmidFinder	100.00	18.42
cBar	34.08	83.27

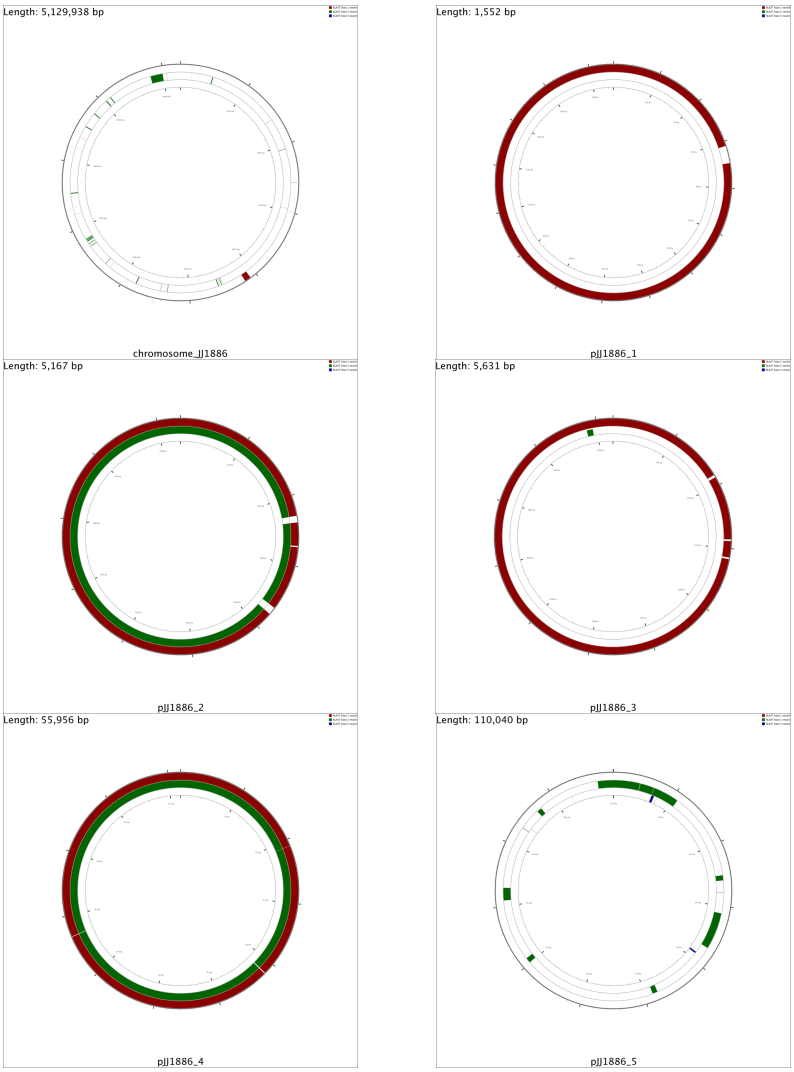

Figure 2. CGViewer representation of the outcome given by PlasmidSPAdes, Recycler and PlasmidFinder against the reference genomes of *E. coli* JJ1886. Minimal alignment 500 bp and cut-off length 70% were selected as parameters by nucleotide local BLAST. PlasmidSPAdes, Recycler and PlasmidFinder are represented by green, red and blue rings respectively.

A.2.2 Citrobacter freundii CFNIH1

PlasmidSPAdes detected a component with a length of 275.6 kbp and composed by 19 contigs (>1 kbp) where the reference plasmid pKEC-a3c was present. In addition, a second component composed by a single contig of 5.4 kbp and an inferred copy number of 14.1 was identified. Recycler was not able to recover the plasmid pKEC-a3c (Figure 3). However, it also extracted the putative plasmid of 5.4 kbp with a coverage ratio of 14.1. We performed a dot-plot of the sequence against itself to observe the presence of circularization signatures at the ends. Furthermore, best blast hit corresponded to "*Klebsiella oxytoca* strain CAV1335 plasmid pCAV-1335-5410, complete sequence" with a length of 5.4 kbp. Annotation made by Prokka identified the presence of mobilization protein MbeC and relaxase MbeA.

The sequence of 5.4 kbp predicted by PlasmidSPAdes and Recycler is the same with a slight difference. Recycler extracted one of the repeat sequences present at the end of the contigs obtaining a final plasmid sequence of 5.409 bp as shown in Figure 4. However, PlasmidSPAdes extracted the plasmid sequences without removing one of the repeats (Figure 5). The previous findings suggested the presence of a complete plasmid sequence of 5.4 kbp which was not previously reported in *C.freundii* CFNIH1.

PlasmidFinder detected the presence of two replication initiator proteins present in pKEC-a3c (Figure 3). The replicon sequences corresponded to the incompatibility groups IncN and IncA.

Table 2. Precision and recall of each program in *C. freundii* CFNIH1.

Program	**Precision (%)**	**Recall (%)**
pSPAdes	98.62	98.71
Recycler	0.00	0.00
PlasmidFinder	100.00	19.42
cBar	68.07	87.48

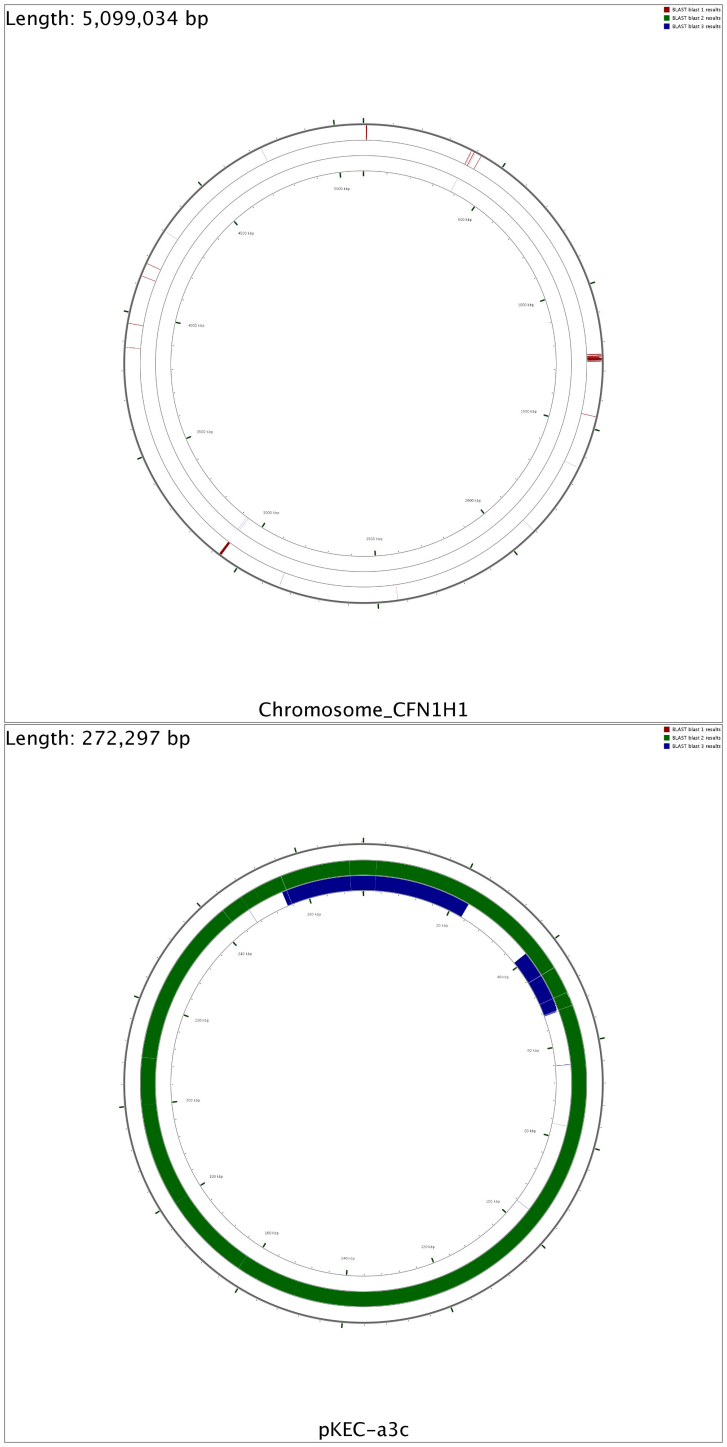

Figure 3. CGViewer representation of the outcome given by PlasmidSPAdes, Recycler and PlasmidFinder against the reference genomes of *C. freundii* CFN1H1. Minimal alignment 500 bp and cutoff length 70% were selected as parameters by nucleotide local BLAST. PlasmidSPAdes Recycler and PlasmidFinder are represented by green, red and blue rings respectively.

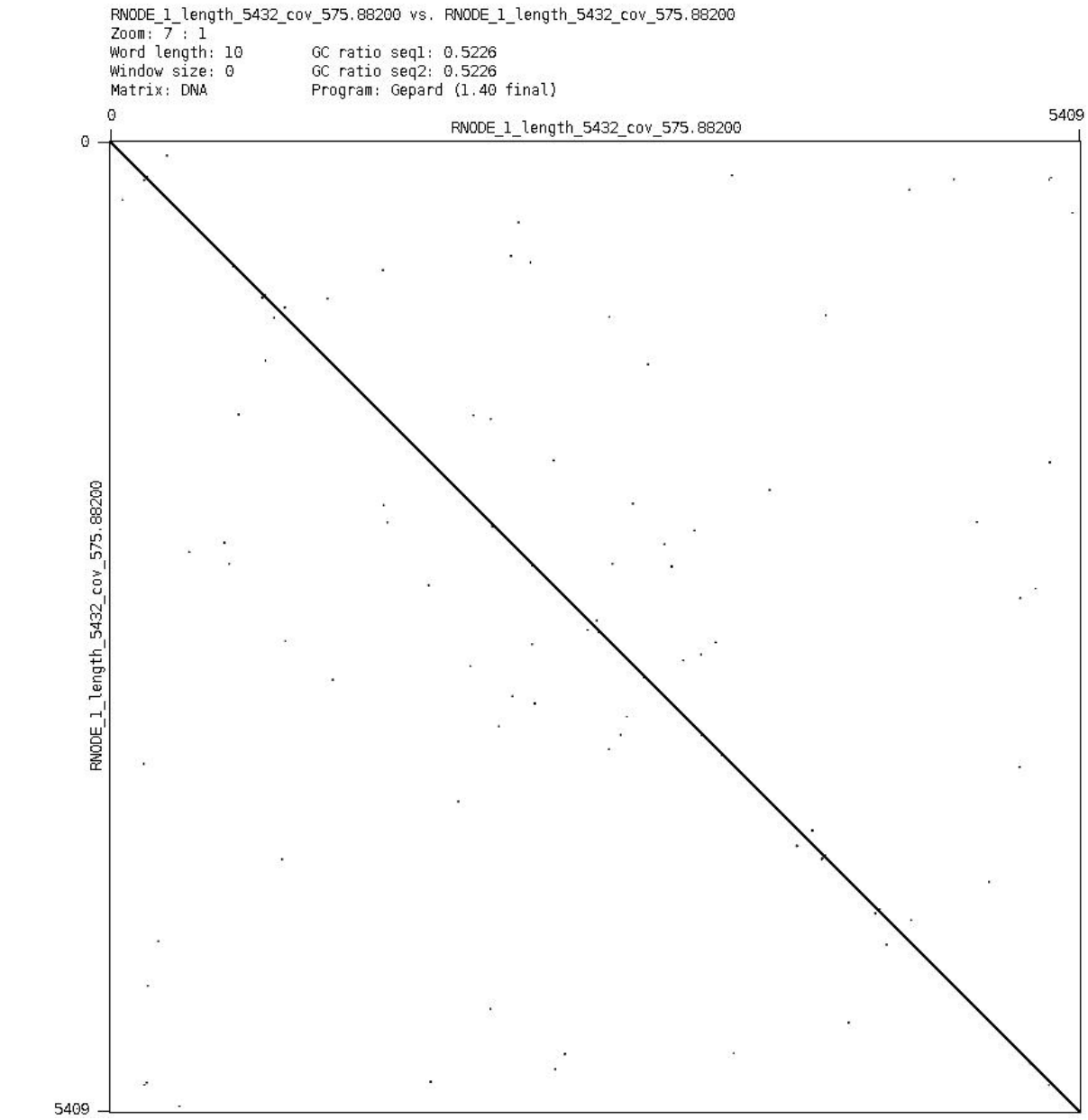

Figure 4. Dot-plot from the new unidentified plasmid from Recycler in *C. freundii* CFNIH1. No repeated sequences are present in both ends of the sequence.

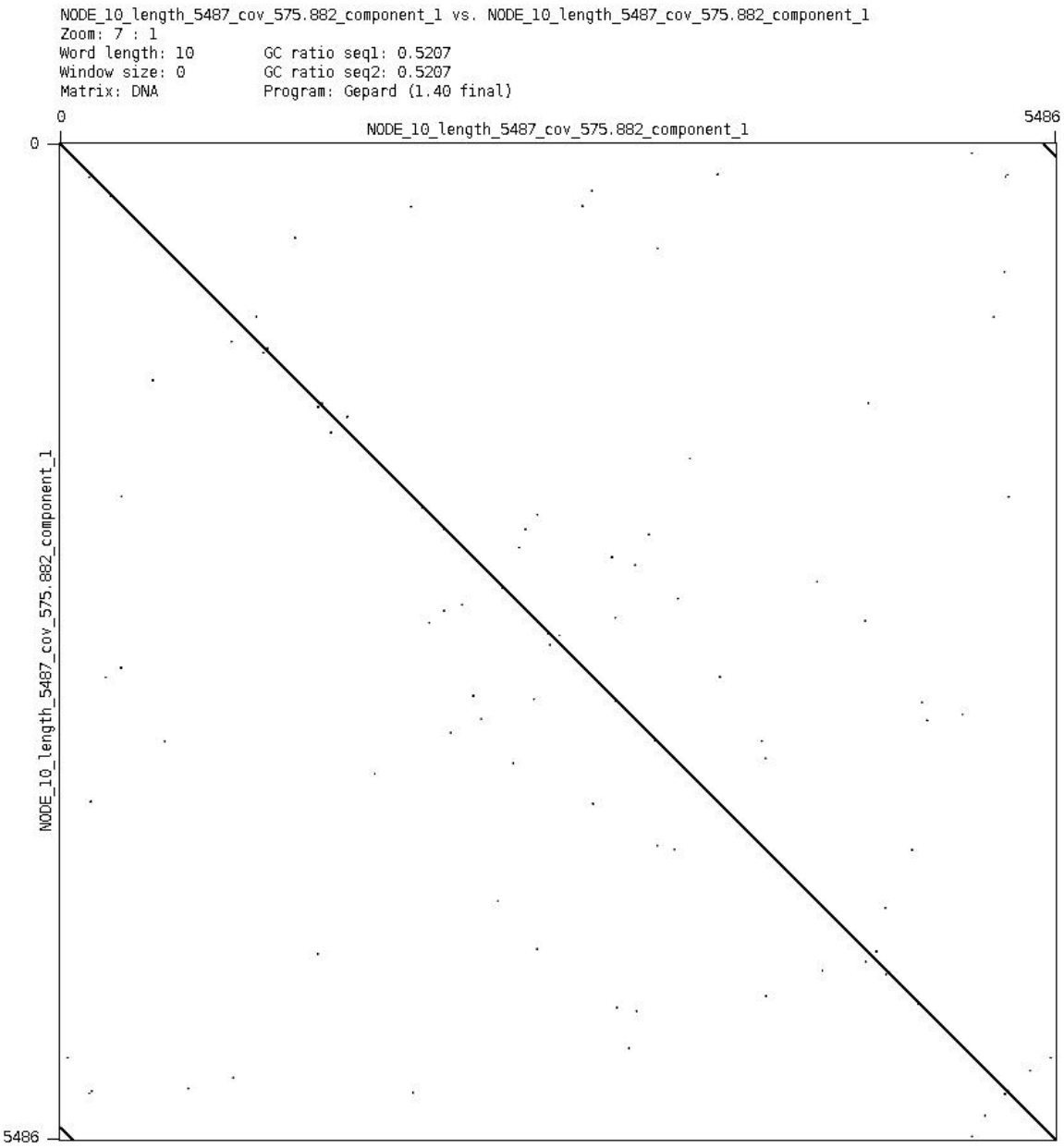

Figure 5. Dot-plot from the new unidentified plasmid from PlasmidSPAdes in *C. freundii* CFNIH1. At both ends of the sequence there is a repeated sequence of 77 bp.

A.2.3 Corynebacterium callunae DSM 20147

PlasmidSPAdes detected two components of 10.4 kbp and 4.2 kbp. A low precision value was obtained because the component of 10.3 kbp was composed by a single contig mapping to the chromosome (Table 3). The component of 4.2 kbp corresponded to the reference plasmid pCC1. Recycler detected exclusively the reference plasmid pCC1 whereas no false positive results were obtained.

cBar obtained a low recall value because only one contig corresponding to pCC2 was correctly identified as plasmid (Table 3). PlasmidFinder was not able to locate any replication initiator sequence in the two reference plasmids present in *C. callunae* DSM 20147. The database of PlasmidFinder was constructed using replicon sequences from the family *Enterobacteriaceae*. Replicon sequences from Gram positive bacteria may differ and may explain the lack of true positive results for this genome.

Table 3. Precision and recall of each program in *C. callunae* DSM 20147.

Program	Precision (%)	Recall (%)
pSPAdes	28.79	4.32
Recycler	100.00	4.23
PlasmidFinder	0.00	0.00
cBar	71.20	9.25

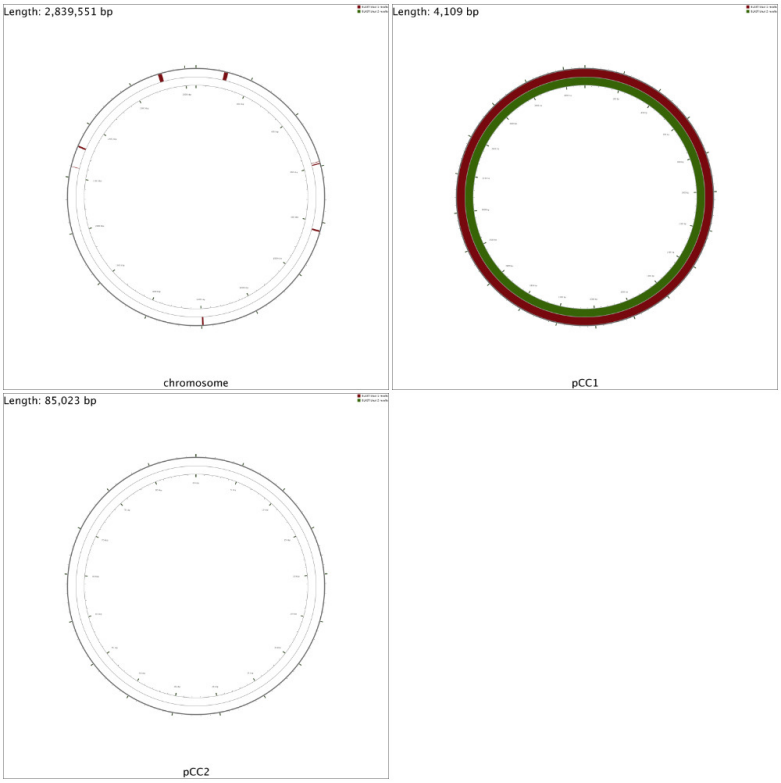

Figure 6. CGViewer representation of the outcome given by PlasmidSPAdes and Recycler against the reference genomes of *C. callunae* DSM 20147. Minimal alignment 500 bp and cutoff length 70% were selected as parameters by nucleotide local BLAST. PlasmidSPAdes and Recycler are represented by green and red rings respectively.

A.2.4 Rhodobacter sphaeroides 2.4.1

PlasmidSPAdes was able to detect a large component of 458 kbp including the five reference plasmids. However, the program was not able to separate the plasmids in different components. PlasmidSPAdes merged them in a single component due to the presence of repeated sequences frustrating the detection of each plasmid as different sub graphs. Visualization of the plasmid graph using Bandage spotted one contig containing a transposase shared in the different physical DNA units (Figure 7). Recycler was only able to detect small fractions from plasmid Ax and plasmid D (Figure 8) whereas lack of false positive results were reported (Table 4). PlasmidFinder did not find plasmid replicon sequences as expected because the bacterial strain belongs to the family *Rhodobacteraceae*.

Table 4. Precision and recall of each program in *R. sphaeroides* 2.4.1.

Program	**Precision**	**Recall**
pSPAdes	100.00	90.77
Recycler	100.00	6.85
PlasmidFinder	0.00	0.00
cBar	59.00	68.51

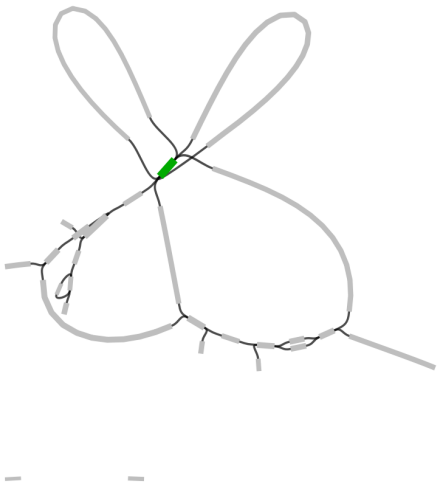

Figure 7. Bandage representation of the assembly graph generated by PlasmidSPAdes in *R. sphaeroides* 2.4.1. Green contig with a length of 1.6 kbp and coverage of 536 was identified as a transposase by BLASTx.

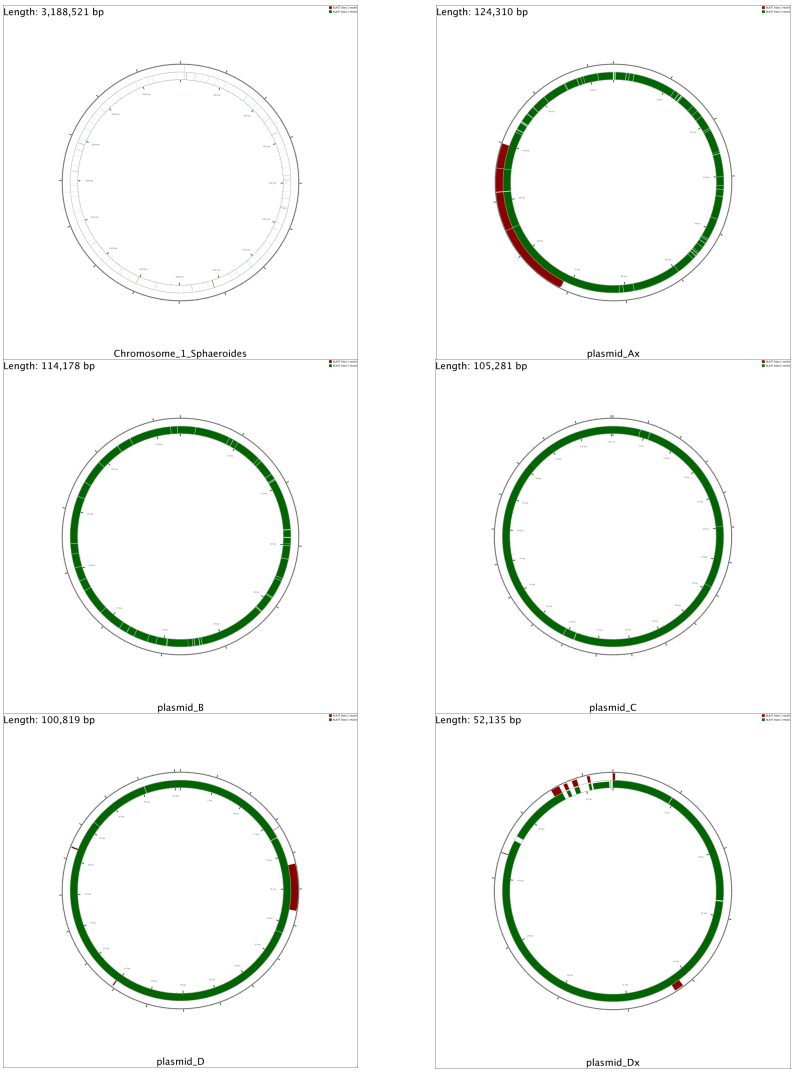

Figure 8. CGViewer representation of the outcome given by PlasmidSPAdes and Recycler against the reference genomes of *R. sphaeroides* 2.4.1. Minimal alignment 500 bp and cutoff length 70% were selected as parameters by nucleotide local BLAST. PlasmidSPAdes and Recycler are represented by green and red rings respectively.

A.2.5 Burkholderia cenocepacia DDS 22E-1

This genome does not contain any reference plasmid but it is composed by three chromosomes with a size of 1.16 Mbp, 3.20 Mbp and 3.66 Mbp (Figure 1). PlasmidSPAdes and Recycler did not detect any plasmid sequence thus the outcomes of both programs corresponded to empty files. Additionally, PlasmidFinder did not find any replicon sequence within the chromosomes of *B.cenocepacia* DDS 22E-1. cBar predicted 1481 contigs (>500 bp) wrongly as plasmid-derived sequences.

A.3 Sequences not mapping to the reference genomes

Genomes described in this section contained contigs predicted by Recycler and PlasmidSPAdes that lacked in the reference assembly of the completed genomes. Only contigs exceeding a minimum length of 1000 bp were further analyzed. Detailed description follows after the tables.

Table 5. Predicted sequences not mapping to the reference assembly. In parenthesis is indicated whether the sequence predicted was an isolated component in the assembly graph.

Program	Strain	Length	Coverage ratio	Best blast hit	Annotation	Circular
pSPAdes	AVNIH1	7241 (Y)	6.3	Plasmid (KT781681)	Antitoxin RelB	Yes
pSPAdes	AVNIH1	1863 (Y)	15.7	Plasmid (LN853312)	-	Yes
Recycler	AVNIH1	7186 (Y)	6.3	Plasmid (KT781681)	Antitoxin RelB	Yes
Recycler	AVNIH1	1808 (Y)	15.7	Plasmid (LN853312)	-	Yes
pSPAdes	700221	12589 (Y)	2.7	Plasmid (AB158402)	-	Yes
pSPAdes	700221	5513 (Y)	26.63	Phage (CP004084)	-	Yes
pSPAdes	700221	3085 (N)	N/A	Plasmid (EU327398)	RepN	No
pSPAdes	700221	1543 (N)	N/A	Plasmid (EU370688)	-	No
pSPAdes	700221	1483 (N)	N/A	CDS (U01917)	RepD	No
Recycler	700221	12534 (Y)	2.7	Plasmid (AB158402)	-	Yes
Recycler	700221	5458 (Y)	26.63	Phage (CP004084)	-	Yes
pSPAdes	AATZP	4294 (Y)	2.44	Plasmid (CP003995)	-	Yes
Recycler	AATZP	4239 (Y)	2.44	Plasmid (CP003995)	-	Yes
pSPAdes	BEST195	5513 (Y)	10.27	Plasmid (CP003995)	-	Yes
Recycler	BEST195	5458 (Y)	10.27	Plasmid (CP003995)	-	Yes
pSPAdes	CAV1392	2572 (Y)	0.12	Plasmid (NC_015515)	-	Yes
Recycler	CAV1392	2517 (Y)	0.12	Plasmid (NC_015515)	-	Yes

Table 6. Predicted sequences not mapping to the reference assembly. In parenthesis is indicated whether the sequence predicted was an isolated component in the assembly graph.

Program	Strain	Length	Coverage ratio	Best blast hit	Annotation	Circular
pSPAdes	KPN223	4.29 (Y)	0.89	Plasmid (EU932690)	-	Yes
pSPAdes	KPN223	4.14 (Y)	1.45	Non significant	Mob A	Yes
Recycler	KPN223	4.23 (Y)	0.89	Plasmid (EU932690)	-	Yes
Recycler	KPN223	4.08 (Y)	1.45	Non significant	Mob A	Yes
Recycler	KPN223	3.62 (Y)	1.28	Plasmid(NZ_CP012489)	-	Yes
pSPAdes	KPN555	4.17 (Y)	0.38	Plasmid (JX238446)	Mob MbeC, Rel MbeA	Yes
pSPAdes	KPN555	4.14 (Y)	1.45	Non significant	Mob A	Yes
Recycler	KPN555	4.23 (Y)	0.89	Plasmid (EU932690)	-	Yes
Recycler	KPN555	4.08 (Y)	1.45	Non significant	Mob A	Yes
Recycler	KPN555	3.62 (Y)	1.28	Plasmid(NZ_CP012489)	-	Yes
pSPAdes	KPNIH27	5.53 (N)	2.43	Non significant	-	Yes
pSPAdes	KPNIH27	3.49 (N)	24.24	Plasmid (CP011643)	-	No
pSPAdes	KPNIH27	2.98 (N)	17.61	Plasmid (CP011619)	-	No
pSPAdes	KPNIH39	5.55 (N)	9.01	Plasmid (NC_019346)	-	Yes
Recycler	KPNIH39	5.66 (N)	9.21	Plasmid (NC_019346)	-	Yes

Table 7. Predicted sequences not mapping to the reference assembly. In parenthesis is indicated whether the sequence predicted was an isolated component in the assembly graph.

Program	Strain	Length	Coverage ratio	Best blast hit	Annotation	Circular
pSPAdes	PMK1	5.69(Y)	26.05	Plasmid (LN854314)	TA HigA-B, Rel MbeA, Mob MbeC	Yes
pSPAdes	PMK1	5.44(Y)	2.02	Scaffold (LL266921)	-	Yes
pSPAdes	PMK1	3.82(Y)	35.02	Plasmid (NC_019077)	-	Yes
Recycler	PMK1	5.69(Y)	26.05	Plasmid (LN854314)	TA HigA-B, Rel MbeA, Mob MbeC	Yes
Recycler	PMK1	5.44(Y)	2.02	Scaffold (LL266921)	-	Yes
Recycler	PMK1	3.82(Y)	35.02	Plasmid (NC_019077)	-	Yes
pSPAdes	CAV1492	1.69(N)	0.07	Plasmid (NC_019346)	-	No
pSPAdes	CAV1741	1.11(N)	0.05	Plasmid (KX276657)	-	No
pSPAdes	ECR091	4.74(Y)	11.83	Plasmid (CP004060)	Rel MbeA, Mob MbeC	Yes
pSPAdes	ECR091	4.38(N)	0.04	Plasmid (CP016526)	-	No
pSPAdes	ECR091	3.61(N)	0.04	Plasmid (CP009856)	-	No
pSPAdes	ECR091	2.57(Y)	22.01	Plasmid (AF014880)	-	Yes
pSPAdes	ECR091	2.47(N)	0.05	Plasmid (CP008899)	-	No
pSPAdes	ECR091	2.31(N)	0.05	Plasmid (CP009856)	-	No
pSPAdes	ECR091	1.90(N)	0.04	Plasmid (CP009856)	-	No
pSPAdes	ECR091	1.89(N)	0.06	Plasmid (CP016526)	-	No
pSPAdes	ECR091	1.41(N)	0.05	Plasmid (CP016526)	-	No
Recycler	ECR091	4.69(Y)	11.83	Plasmid (CP004060)	Rel MbeA, Mob MbeC	Yes
pSPAdes	ECNIH3	2.57 (Y)	29.70	Plasmid (AF014880)	-	Yes
Recycler	ECNIH3	2.58 (Y)	32.16	Plasmid (AF014880)	-	Yes

A.3.1 Aeromonas veronii strain AVNIH1

PlasmidSPAdes identified two components formed by a single contig not mapping to the reference assembly of *A. veronii* strain AVNIH1. In addition, Recycler also identified the same putative plasmids and it extracted one of the repeated sequences present at both ends of the contigs (Figure 9). The largest sequence had as best blast hit "*Aeromonas salmonicida* subsp. salmonicida strain JF2507 plasmid pAsal1D, complete sequence" with a length of 9.1 kbp. The other sequence corresponded to "Uncultured prokaryote from Rat gut metagenome metamobilome, isolate RGRH0694" with a length of 1.8 kbp. The presence of circularization signatures and a similar best blast hit length suggested the presence of two small complete plasmids previously not reported for this strain with a length of 7113 bp and 1735 bp.

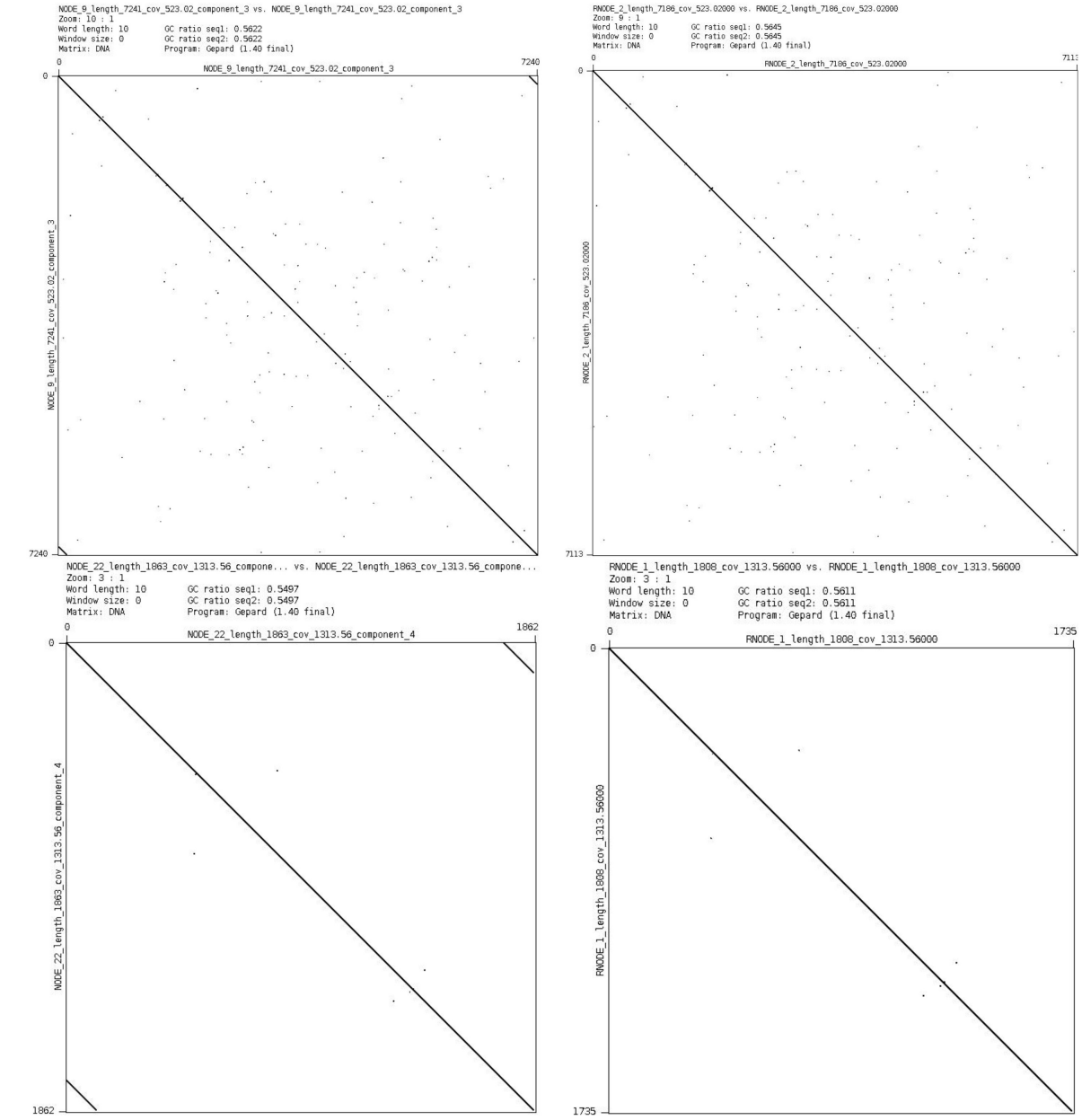

Figure 9. Dot-plot from the sequences not mapping to the reference genomes *A. veronii* strain AVNIH1. On the left side, the sequences not mapping to the reference assembly from PlasmidSPAdes are represented. It is possible to observe repeated sequences at both ends of the contig. On the right side the results from Recycler are represented where one of the repeated sequences has been removed it.

A.3.2 Enterococcus faecium strain ATCC 700221

Both programs identified the presence of two sequences not mapping to the reference assembly with a length of 12.5 kbp and 5.4 kbp. The largest sequence had as best blast hit "*E. faecium* plasmid p200B" with a length of 12.5 kbp. The sequence of 5.4 kbp had as best blast hit "Enterobacteria phage phiX174, complete genome" with the same length. Additionally, in both cases there was presence of circularization signatures (Figure 10). These findings suggest the identification of a plasmid not present in the reference assembly with a length of 12,461 bp.

In addition, PlasmidSPAdes identified three sequences, included in the largest component (length of 1.22 Mbp), with a length of 3.08 kbp, 1.54 kbp and 1.48 kbp. The sequences had a blast hit corresponding to plasmids with a similar size and the presence of replicon sequences (Table 5). However, the absence of circularization signatures at the end of the sequences does not allow to confirm the completeness of these sequences not mapping to the reference assembly.

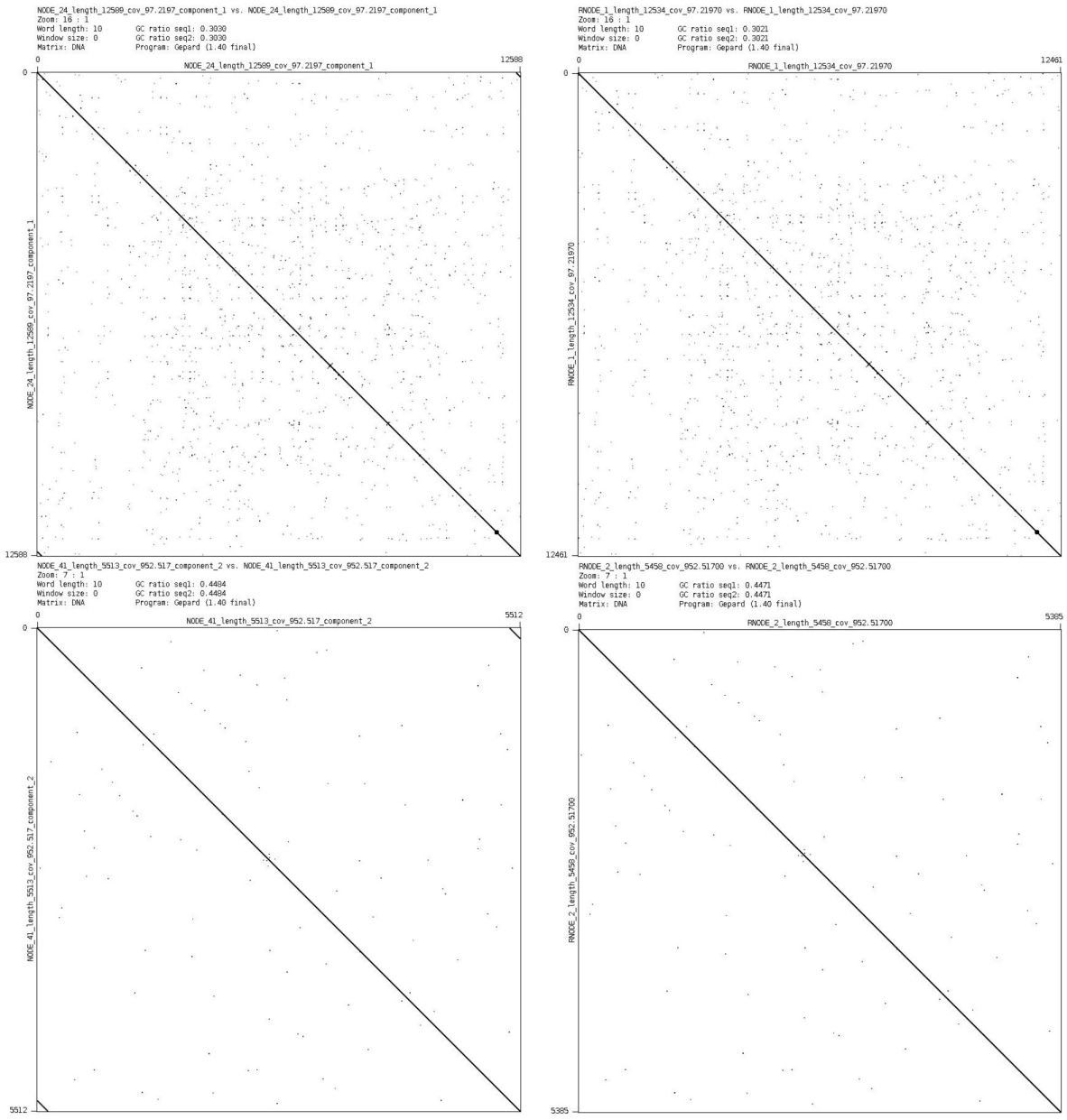

Figure 10. Dot-plot from the sequences not mapping to the reference genomes *E. faecium* strain ATCC 700221. On the left side, the sequences not mapping to the reference assembly from PlasmidSPAdes are represented. It is possible to observe repeated sequences at both ends of the contig. On the rigth side the results from Recycler are represented where the complete plasmid sequence was correctly identified.

A.3.3 Klebsiella pneumoniae strain AATZP

PlasmidSPAdes identified an isolated component formed by one contig of 4.29 kbp. Additionally, Recycler did not identify any reference plasmid present in *K. pneumoniae* strain AATZP but it also detected the same unidentified plasmid. Best blast hit corresponded to: "*Klebsiella pneumoniae* subsp. pneumoniae Kp13 plasmid pKP13c, complete sequence" with a length of 5.06 kbp. A similar blast hit length and the presence of circularization signatures at both ends of the sequence suggested the identification of a small cryptic plasmid with a length of 4166 bp not present in the reference assembly of *K. pneumoniae* strain AATZP (Figure 11).

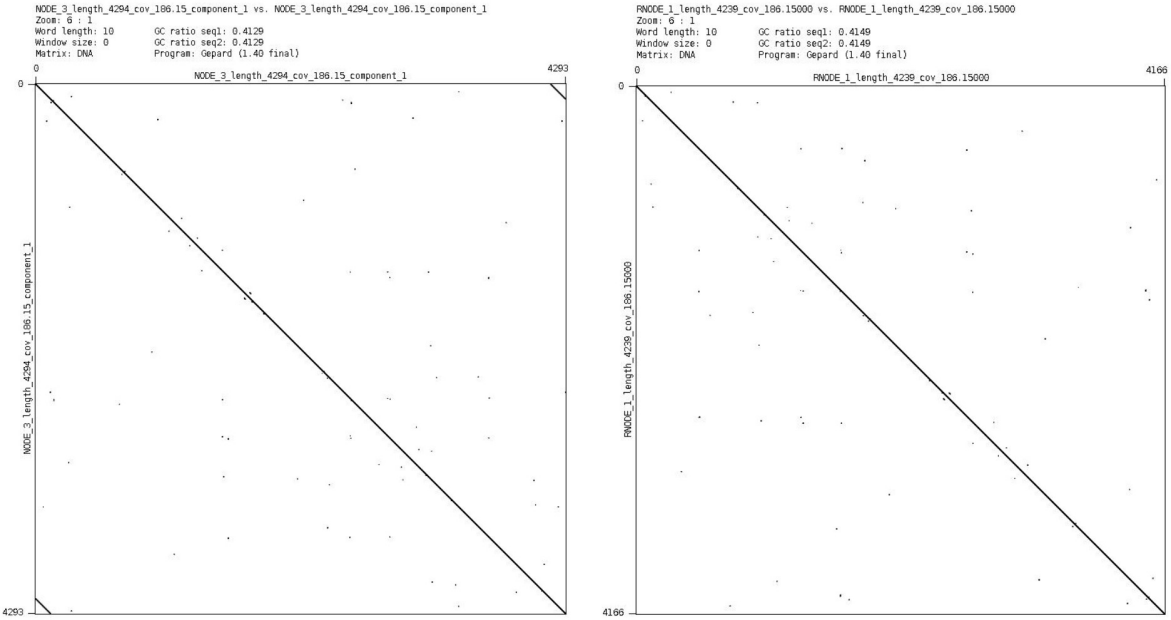

Figure 11. Dot-plot from the sequence not mapping to the reference genomes *K. pneumoniae* strain AATZP. On the left side, the sequence not mapping to the reference assembly from PlasmidSPAdes is represented. Repeated sequences at both ends of the contig are present. On the rigth side the results from Recycler are represented.

A.3.4 Bacillus subtilis subsp. natto BEST195

PlasmidSPAdes and Recycler identified a sequence not mapping to the reference genomes with a length of 5.5 kbp, present in a coverage ratio of 10 and with circularization signatures (Figure 12). It may correspond to the plasmid pBEST195L which sequence has been reported but it has not been deposited yet to the nr database NCBI [28].

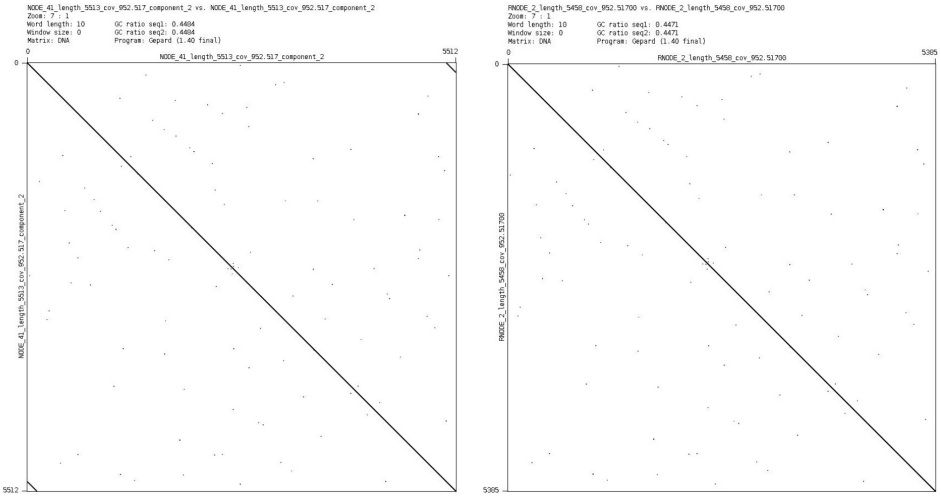

Figure 12. Dot-plot from the sequence not mapping to the reference genomes *B. subtilis* subsp. natto BEST195. On the left side, the sequence not mapping to the reference assembly from PlasmidSPAdes is represented. Repeated sequences at both ends of the contig are present. On the right side the results from Recycler are represented.

A.3.5 Klebsiella pneumoniae strain CAV1392

PlasmidSPAdes identified an isolated component composed by a single contig of 2.57 kbp with a coverage ratio of 0.12. Additionally, Recycler identified the same sequence with the same coverage. Best blast hit corresponded to "*Enterobacter sp.* W001 plasmid pR23, complete sequence" with a length of 10.49 kbp. The presence of circularization signatures at the end suggested the completeness of the plasmid (Figure 13). However, small cryptic plasmids are usually present in a intermediate to high copy number. For this reason, further studies and validations may be necessary if its presence is the result of contamination.

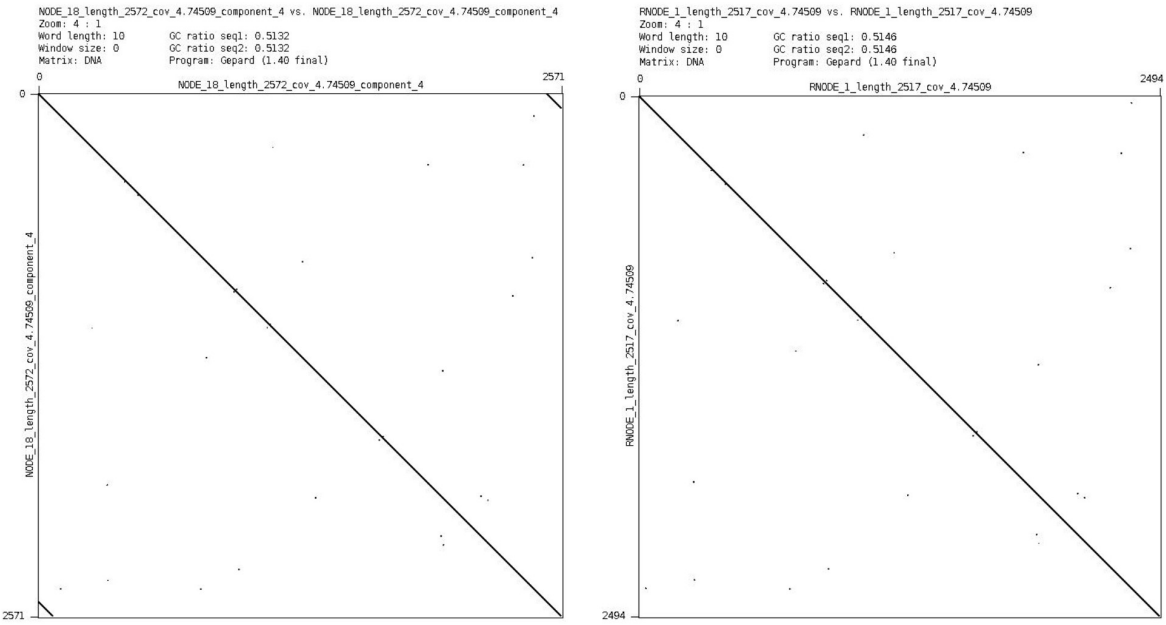

Figure 13. Dot-plot from the sequence not mapping to the reference genomes *K. pneumoniae* strain CAV1392. On the left side, the sequence from PlasmidSPAdes is represented. Repeated sequences are present at both ends of the contig. On the rigth side the sequence not mapping to the reference assembly from Recycler is represented.

A.3.6 Klebsiella pneumoniae strain Kpn223

PlasmidSPAdes identified two isolated components formed by a single contig with a length of 4.29 kbp and 4.14 kbp. Recycler identified the same sequences as plasmids and it also detected an extra sequence of 3.62 kbp. The sequence of 4.29 kbp had as best blast hit "*Klebsiella pneumoniae* strain 0773 plasmid pKpn114, complete sequence" with a length of 4.21 kbp. The sequence of 4.14 kbp did not have any significant blast hit even though Prokka detected the presence of a Mobilization protein A and there was indication of a circularization signatures (Table 6). In addition, the sequence identified by Recycler with a length of 3.62 kbp had a best blast hit with "*Enterobacter sp.* FY-07 plasmid pAKI40B, complete sequence". The last findings suggested the presence of three small cryptic plasmids which were not previously reported in *K. pneumoniae* strain Kpn223.

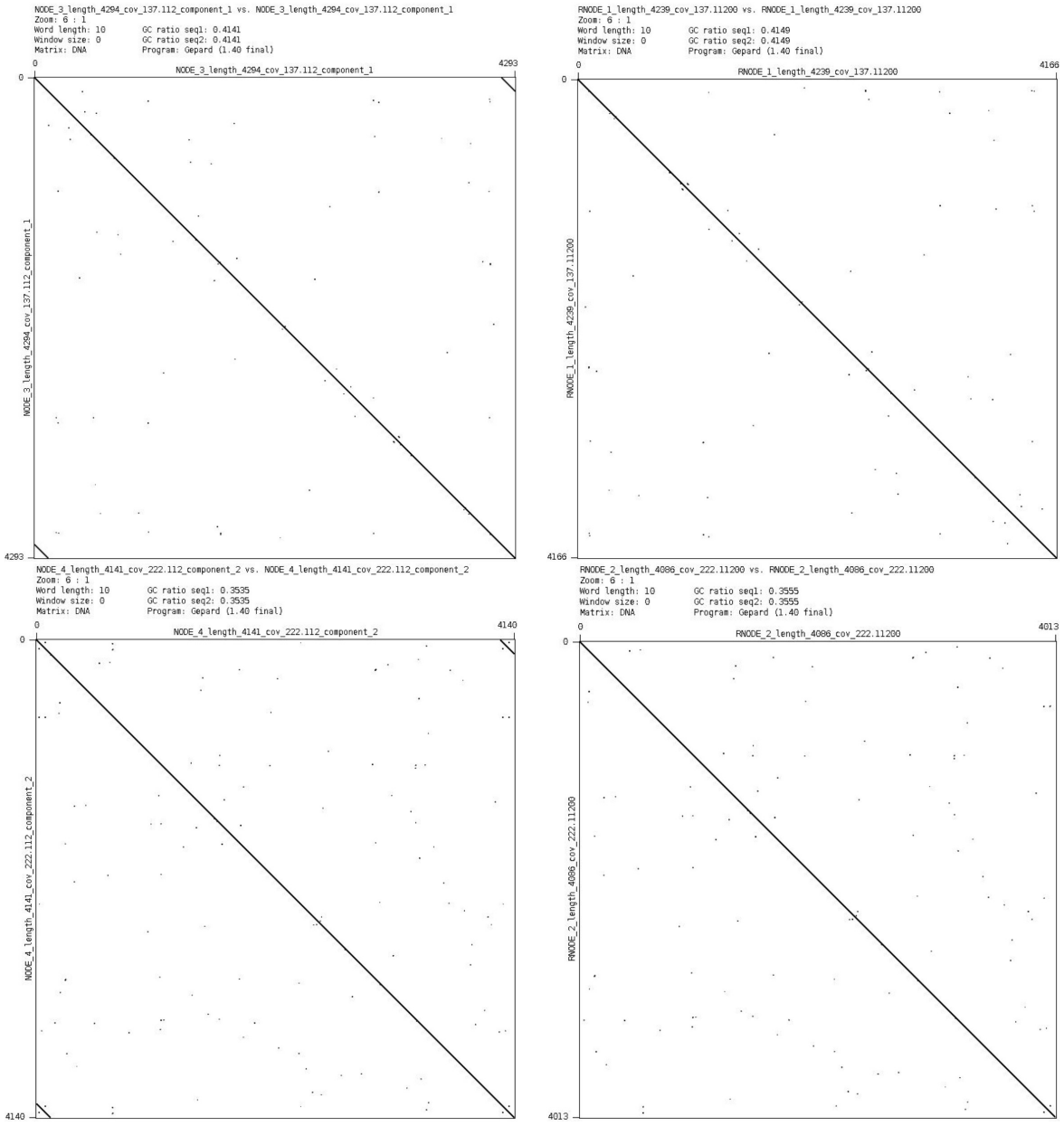

Figure 14. Dot-plot from the sequences not mapping to the reference genomes *Klebsiella pneumoniae* strain Kpn223. On the left side, the unidentified sequence from PlasmidSPAdes is represented. It is possible to observe a repeated sequences at both ends of the contig. On the rigth side the results from Recycler are represented.

A.3.7 Klebsiella pneumoniae strain Kpn555

PlasmidSPAdes and Recycler identified the same sequences not mapping to the reference assembly. The sequence of 4.1 kbp had as best blast hit "*Escherichia coli* strain EC19 plasmid pEC19-1 hypothetical proteins, MobA, MobB, and MobC genes, complete cds" with a length of 4.86 kbp. Sequence annotation and circularization signatures indicated the presence of a plasmid (Table 6) present in a low-copy number.

The sequence with a length of 3.55 kbp had as best blast hit "*Klebsiella pneumoniae* subsp. pneumoniae MGH 78578 plasmid pKPN7, complete sequence" with a length of 3.78 kbp. Prokka anotation and circularization signatures (Figure 15) suggested the presence of a plasmid with a length of 3487 bp which was previously not reported in *Klebsiella pneumoniaea* Kpn555.

Finally, two sequences with a length of 3.00 kbp and 2.92 kbp were also reported. In both cases, the best blast hit corresponded to "Uncultured bacterium extrachromosomal DNA RGI00802" with a length of 2.80 kbp. Contig annotation and circularization signatures suggested that these could be two small plasmids.

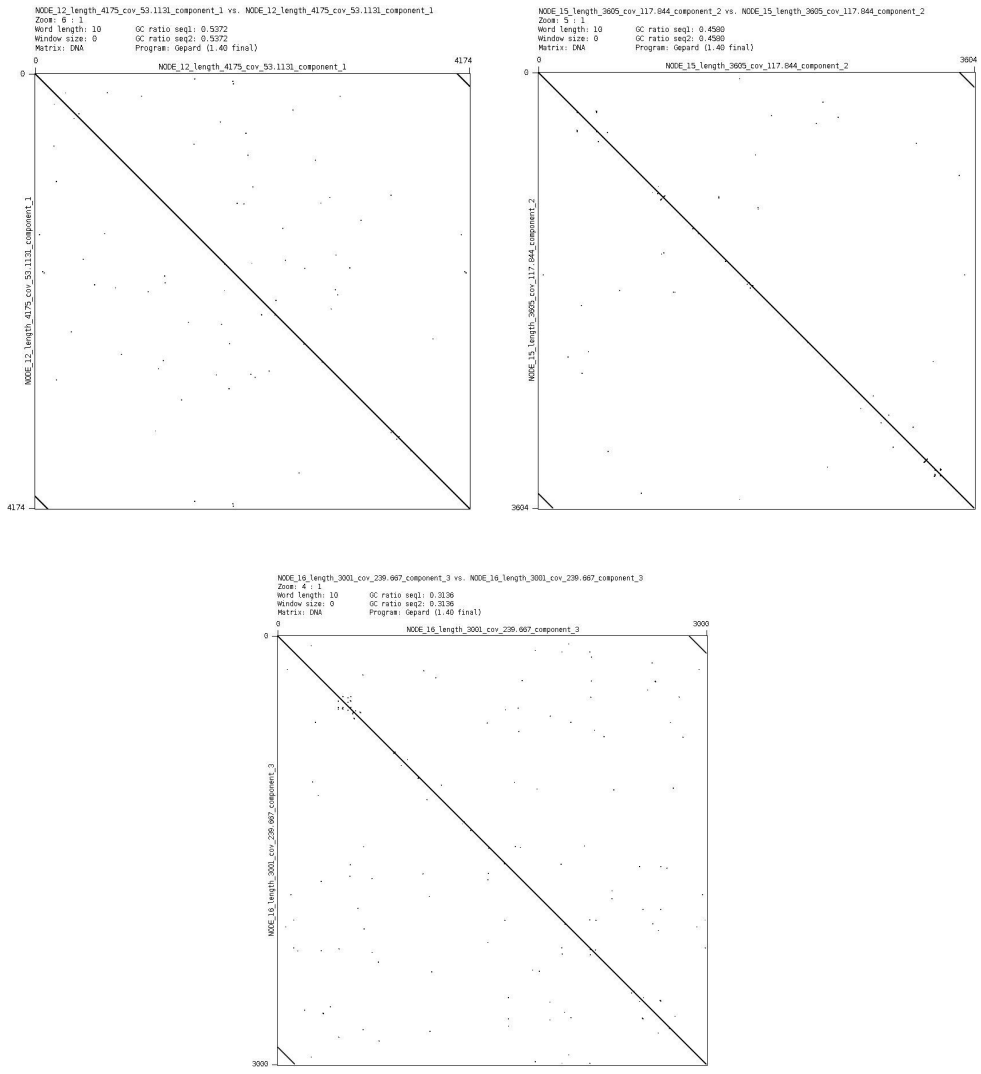

Figure 15. Dot-plot from the sequences not mapping to the reference genomes *Klebsiella pneumoniae* strain Kpn555. Only the sequences predicted by PlasmidSPAdes with a length of 4.17 kbp, 3.60 kbp and 3.00 kbp are represented.

A.3.8 Klebsiella pneumoniae strain KPNIH27

PlasmidSPAdes identified three contigs with a length of 5.53 kbp, 3.49 kbp and 2.98 kbp. The sequences were not detected as isolated components and they were present in the largest component of the assembly graph with a length of 958.61 kbp (Table 6).

The sequence of 5.53 kbp had as best blast hit "*Klebsiella pneumoniae* subsp. pneumoniae KPNIH27, complete genome" which corresponds to the chromosome of this genome. Query coverage of 17% indicated why this contig was not mapped against the chromosome using Quast. In addition, the same repetitive sequence was present at the end of the sequence (Figure 16. Further studies would be necessary to confirm if the sequence is integrated in the chromosome or it is an extrachromosomal plasmid with chromosome sequences integrated.

The other two sequences had a best blast hit corresponding to a plasmid and were present in a high copy number (Table 6). However, no circularization signatures were present and it is not possible to confirm the completeness of these sequences.

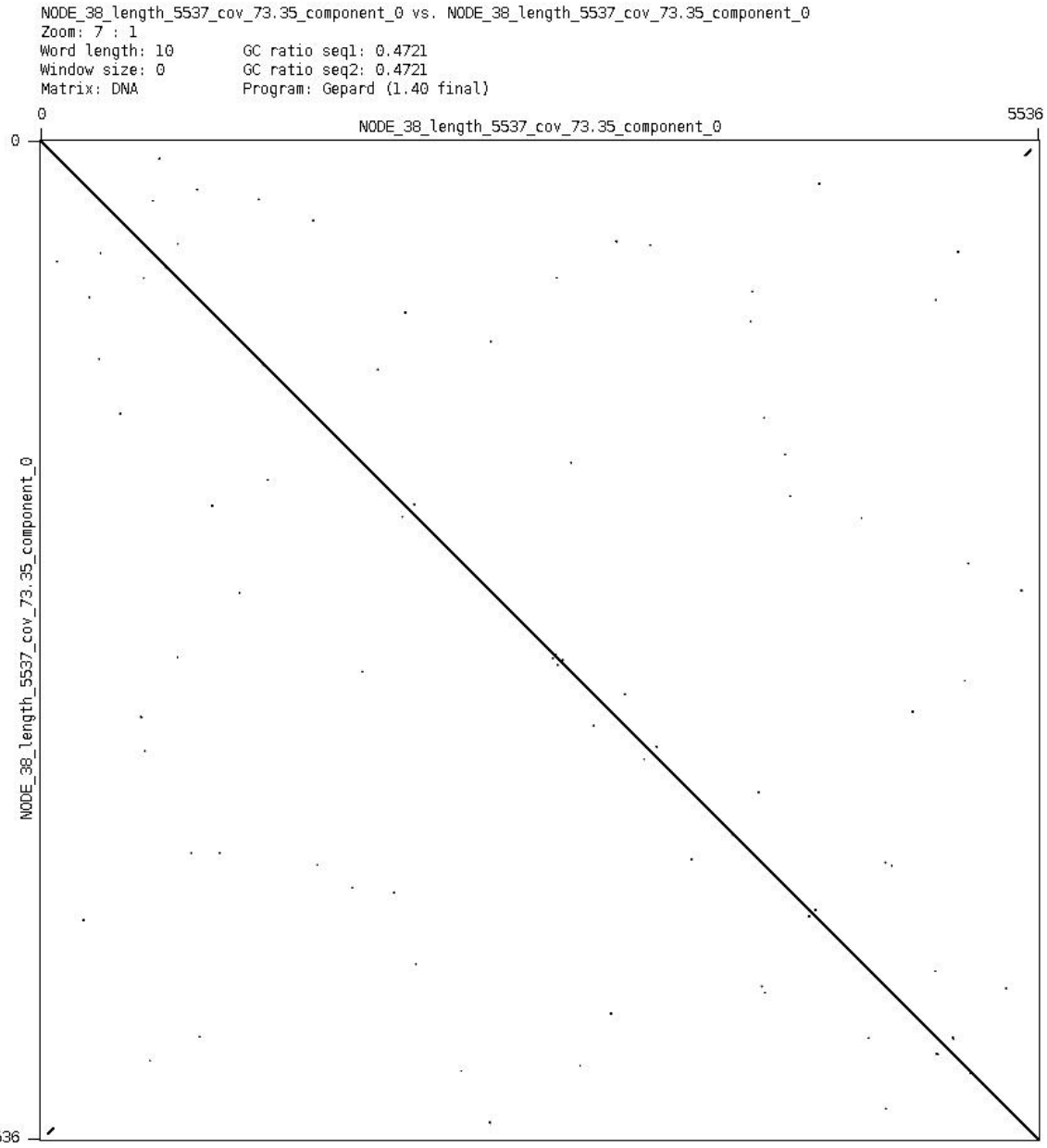

Figure 16. Dot-plot from the contig of 5.53 kbp not mapping to the reference genomes *Klebsiella pneumoniae* strain KPNIH27. It is possible to observe a repeated sequence at both ends of the contig

A.3.9 Klebsiella pneumoniae strain KPNIH39

PlasmidSPAdes and Recycler identified the same sequence with a length of 5.6 kbp. Best blast hit corresponded to "*Enterobacter cloacae* plasmid pNE1280, complete sequence". Presence of circularization signatures and a copy number of 9.0 (as inferred by contig coverage) indicated the presence of a small cryptic plasmid with a length of 5549 bp.

A.3.10 Klebsiella pneumoniae strain PMK1

PlasmidSPades and Recycler identified the same three sequences not present in the reference assembly.

Contig annotation spotted the presence of plasmid-genes related in the sequence of 5.6 kbp. Best blast hit corresponded to "Uncultured prokaryote from Rat gut metagenome metamobilome, plasmid pRGRH1815" with a length of 7.10 kbp. The presence of circularization signatures and a high-copy number suggested the characterization of a plasmid with a length of 5639 bp.

In addition, the sequence with a length of 3.8 kbp presented circularization signatures and a best blast hit corresponding to "*Escherichia coli* strain NCTC 9034 plasmid pEC34A, complete sequence". Dot-plot allowed to detect the sequence was identical than the one reported by PlasmidSPAdes and Recycler. Additionally, circularization signatures were present indicating the completeness of the plasmid. The previous findings suggested the presence of a plasmid with a length of 3770 bp.

Finally the sequence with a length of 5.4 kbp had as best blast hit "*Echinostoma caproni* genome assembly E_caproni_Egypt, scaffold ECPE_contig0001929". Several blast hit results with a similar bit-score indicated the presence of a phage. This may explain the presence of circularization signatures at the ends of the sequence.