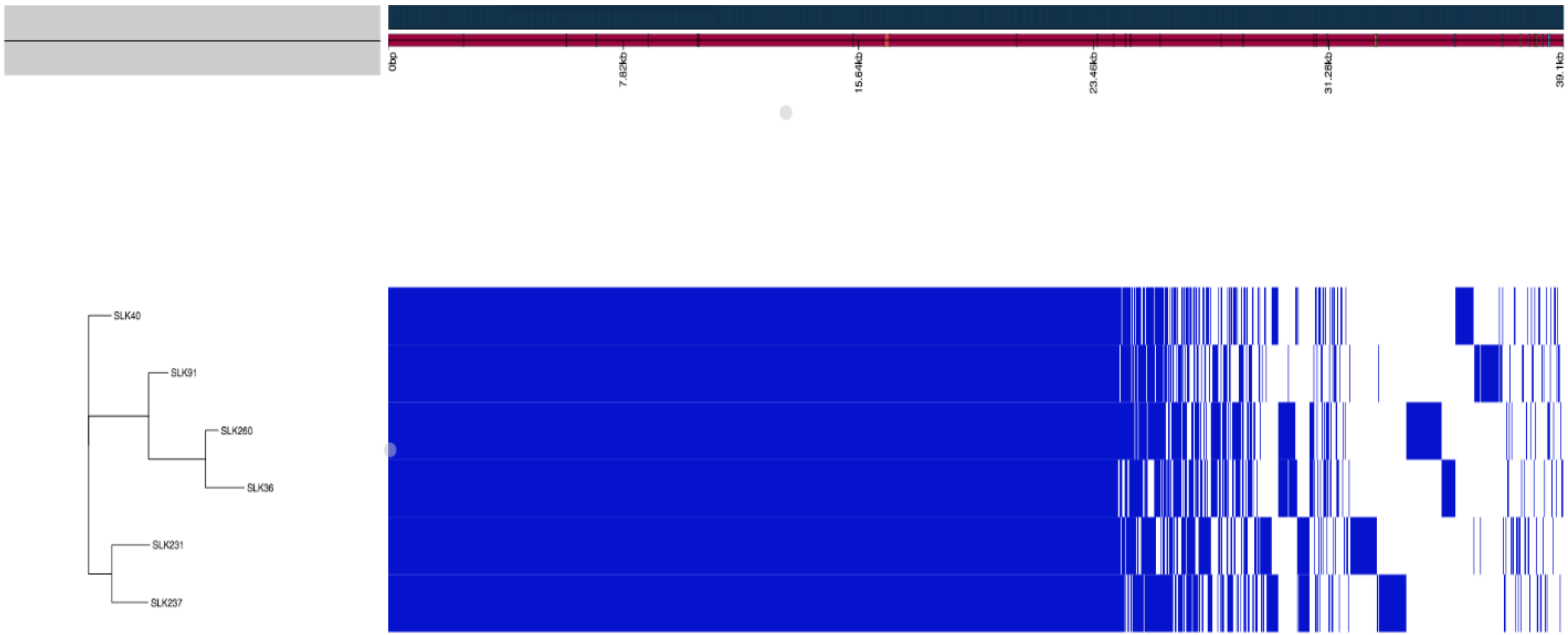
Supplementary Table S1. Quality of assembly by DFAST

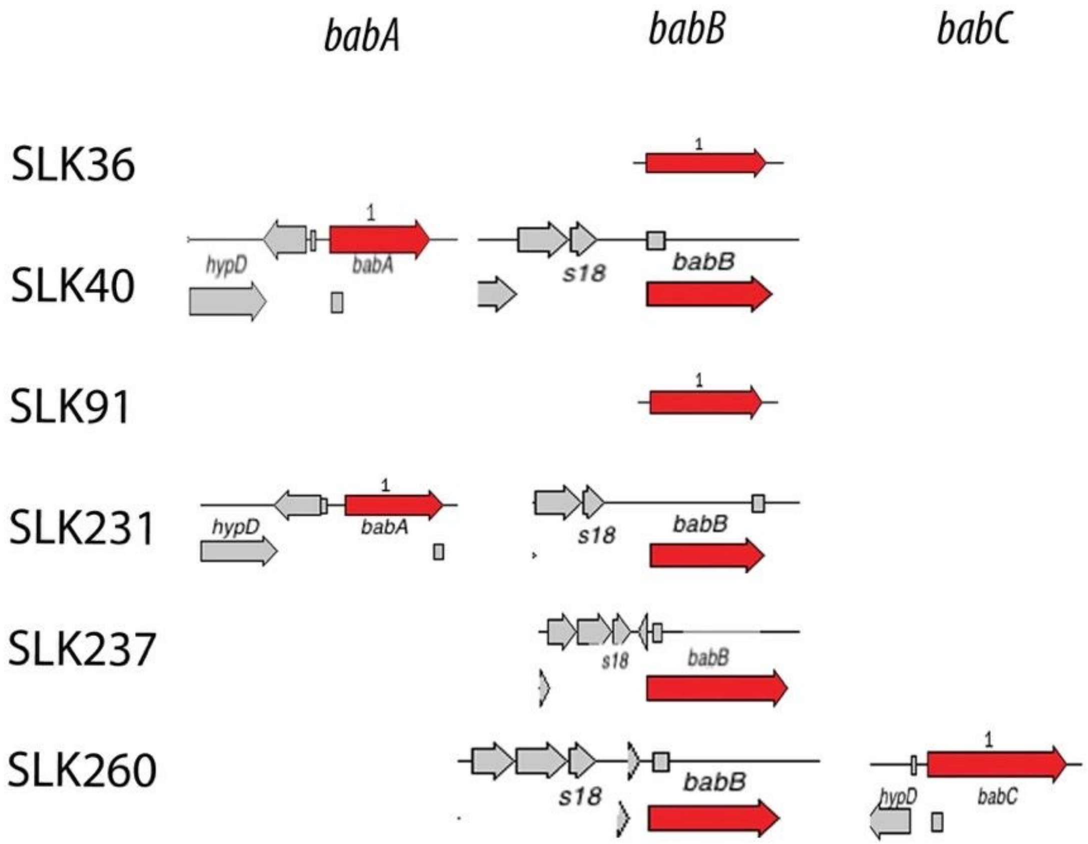
	SLK91	SLK40	SLK36	SLK260	SLK237	SLK2
Total Sequence Length (bp)	1552574	1534541	1585718	1625743	1584937	15721
Number of Sequences	45	29	48	39	39	48
Longest Sequences (bp)	229567	303249	199185	232082	350061	24543
N50 (bp)	85701	137496	60855	80865	93038	11831
Gap Ratio (%)	0	0	0	0	0	0
GCcontent (%)	39.1	39.2	38.9	38.8	39.1	39.2
Number of CDSs	1440	1442	1453	1509	1495	1481
Average Protein Length	321.7	321.5	324.3	323.3	318.1	315.7
Coding Ratio (%)	89.5	90.6	89.1	90	90	89.2
Number of rRNAs	2	2	2	2	2	2
Number of tRNAs	36	36	36	36	36	36
Number of CRISPRs	1	0	1	0	1	1

Supplementary Table S2. The presence of phage among Sri Lanka isolates

Strains	Phage	Size	Type	GC Percentage	Highest Similarity
SLK36	1	18.9Kb	incomplete	41.55%	PHAGE_Bacill_G_NC_023719(2)
SLK40	None				
SLK91	1	19Kb	incomplete	41.34%	PHAGE_Bacill_AR9_NC_031039
SLK231	1	18.9Kb	incomplete	41.55%	PHAGE_Bacill_G_NC_023719(2)
SLK237	1	14.9Kb	intact	36.10%	PHAGE_Helico_KHP30_NC_0199
	2	18.9Kb	incomplete	41.57%	PHAGE_Bacill_PBS1_NC_043027
Slk260	1	19Kb	incomplete	41.35%	PHAGE_Bacill_AR9_NC_031039



Supplementary Figure S1. Visualization of the core genome and shell genome of Sri Lanka strains. These results showed the Sri Lanka genomes and 26695

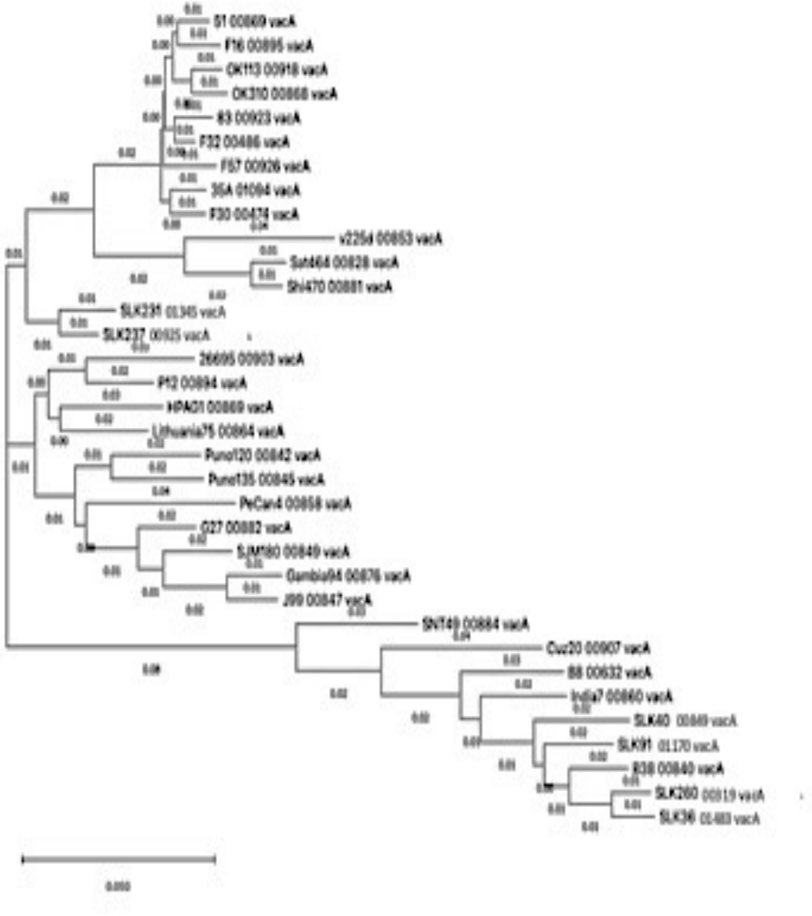


Supplementary Figure S2. Determination of *babA, babB, and babC* locus. This figure showed the position of *bab* gene relative to other genes to determine locus. *bab* gene is located in locus A if located downstram to *hypD* , locus B if the gene is located downstream to *s18*.

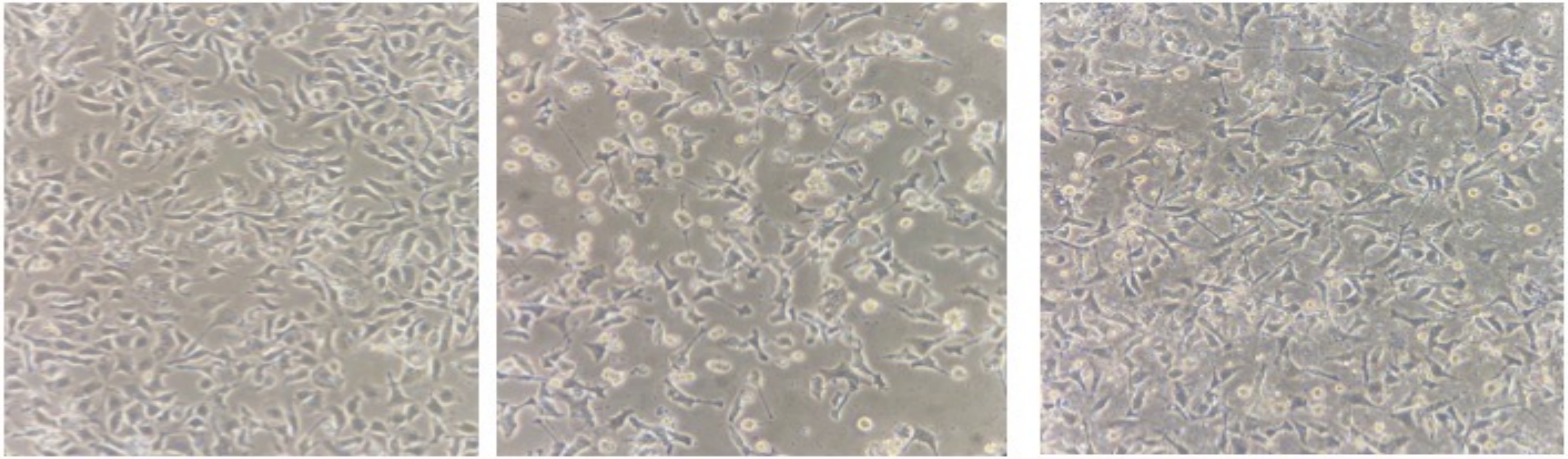
cagA Alignment

```
26695  KEQLAQQAQKNEDFNTGKNSELYQSVKNSVNKTLVGNGLSGIEATALAKNFSDIKKELNEKFK
SLK237 KEQLAQQAQKNESFNVGKKSEIYQSVKNGVNGTLVGNGLSKAEATTLSKNFSDIKKELNAKLG
SLK231 KEQLVQQAQKNESFNVGKKSEIYQSVKNGVNGTLVGNGLSKAEATTLSKNFSDIRKELSEKLG

26695  NFNNNN-NGLKNSTEPIYAKVNKKKTGQVASPEEPIYTQVAKKVNAKIDRLNQIASGLGGVGQ
SLK237 NFNNNNNNGLKN--EPIYAKVNKKKAGQAASLEEPIYAQVAKKVNAKIDRLNQIASGLGGVGQ
SLK231 NFNNNN-NGLKN---------------------EPIYAQVAKKVNAKIDRLNQIARGLGGVGQ

26695  AAGFPLKRHD-----------------------------------------------------
SLK237 AAGFPLKRHD-----------------------------------------------------
SLK231 AAGFPLKKHDKVDDLSKVGLSASPEPIYATIDDLGGPFPLKKHDKVDDLSKVGLSASPEPIYA

26695  ---------------KVDDLSKVGLSASPEPIYATIDDLGGPFPLKRHDKVDDLSKVGRSRNQ
SLK237 ---------------KVDDLSKVGRSVSPEPIYATIDDLGGPFPLKRHDKVDDLSKVGLSRNR
SLK231 TIDDLGGPFPLKRHDKVDDLSKVGLSASPEPIYATIDDLGGPFPLKRHDKVDDLSKVGLSRNQ
```

Supplementary Figure S3. Sequence alignment of *cagA* 26695 in the EPIYA region and CM motifs



Supplementary Figure S4. The phylogenetic tree of *vacA* compared to the *vacA* from the reference genomes. Sri Lankastrains possessing s1m1 genotype represented in a distinct branch.



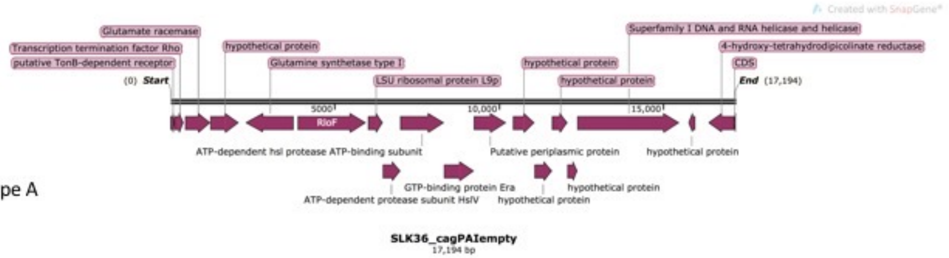
SLK36

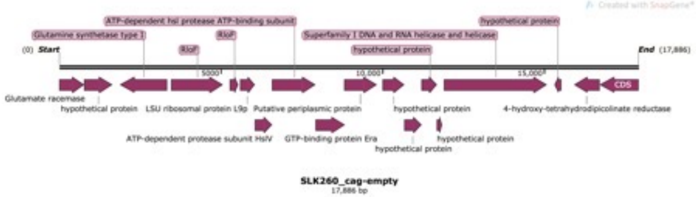
SLK231 (BCCC)

SLK237 (ABC)

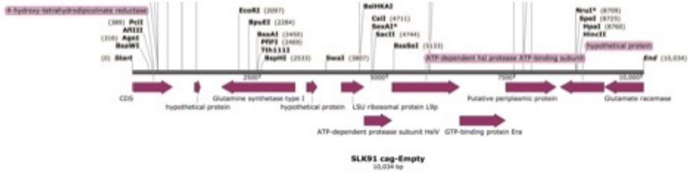
Supplementary Figure S5. The hummingbird phenotype of AGS cells infected with Sri Lanka strains (40X magnification). Each of figure represent the CagA negative, CagA with BCCC genotype and CagA with ABC genotype.

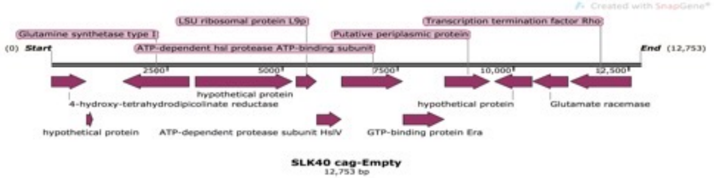
Pre type A



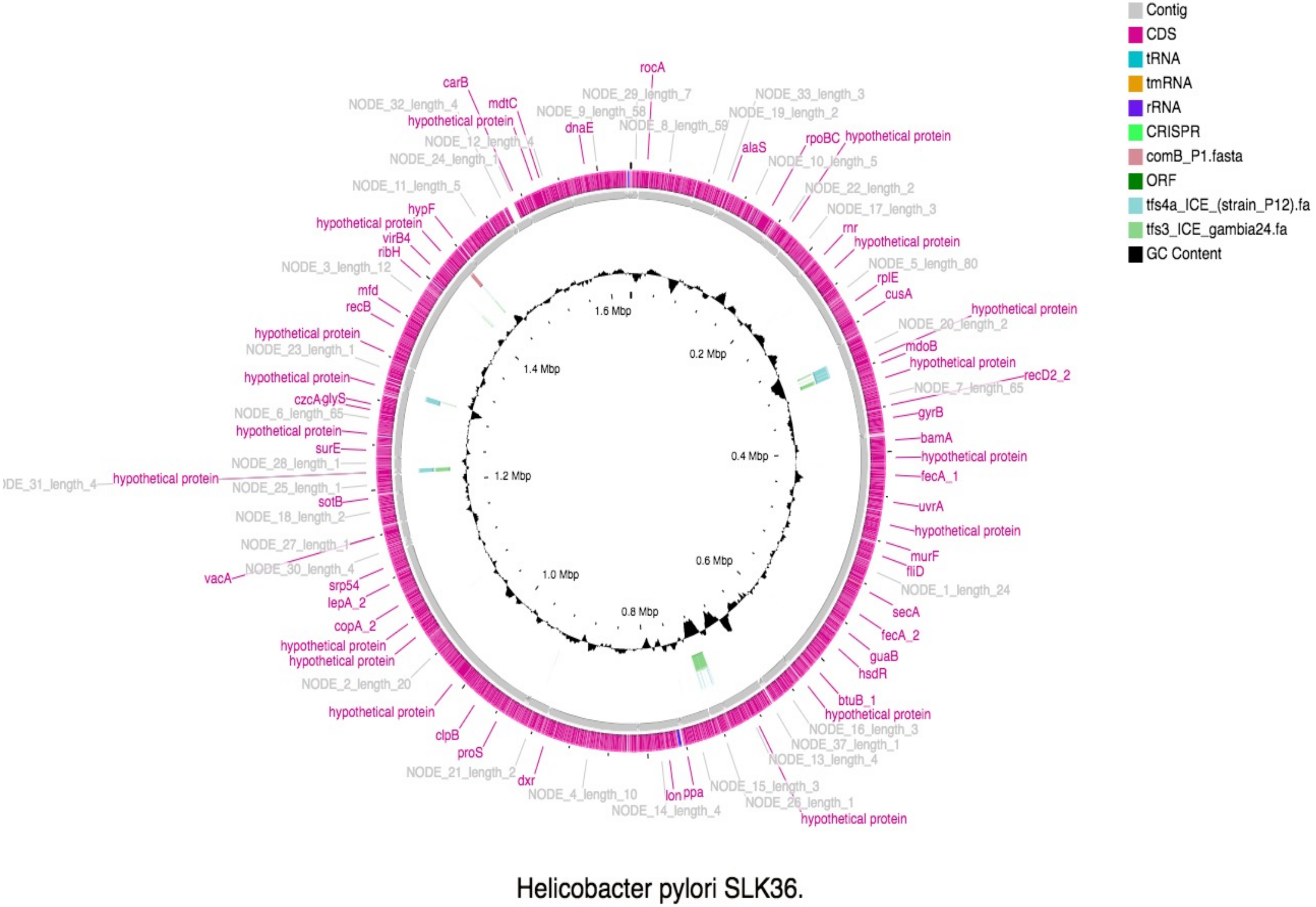


Pre type B

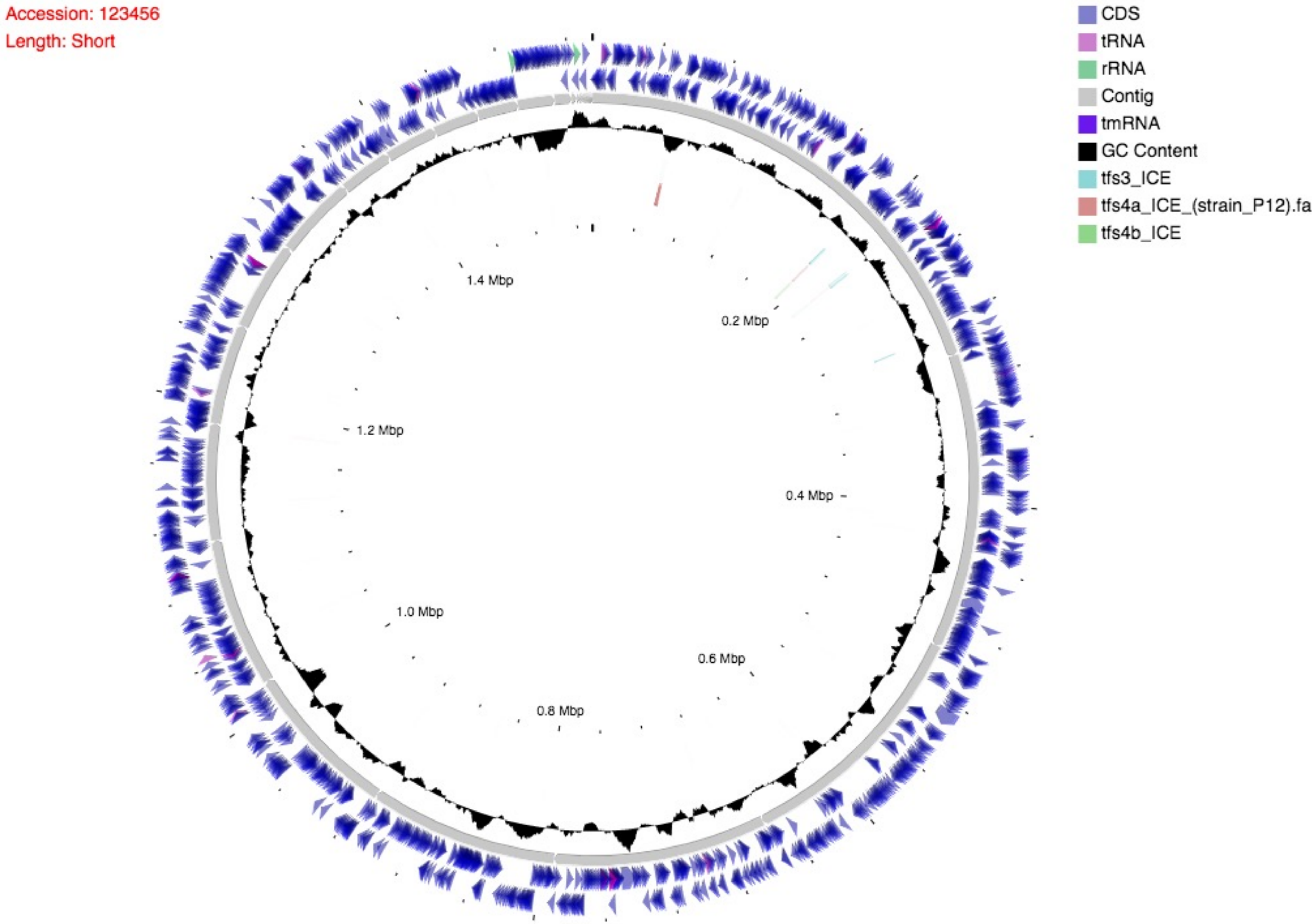




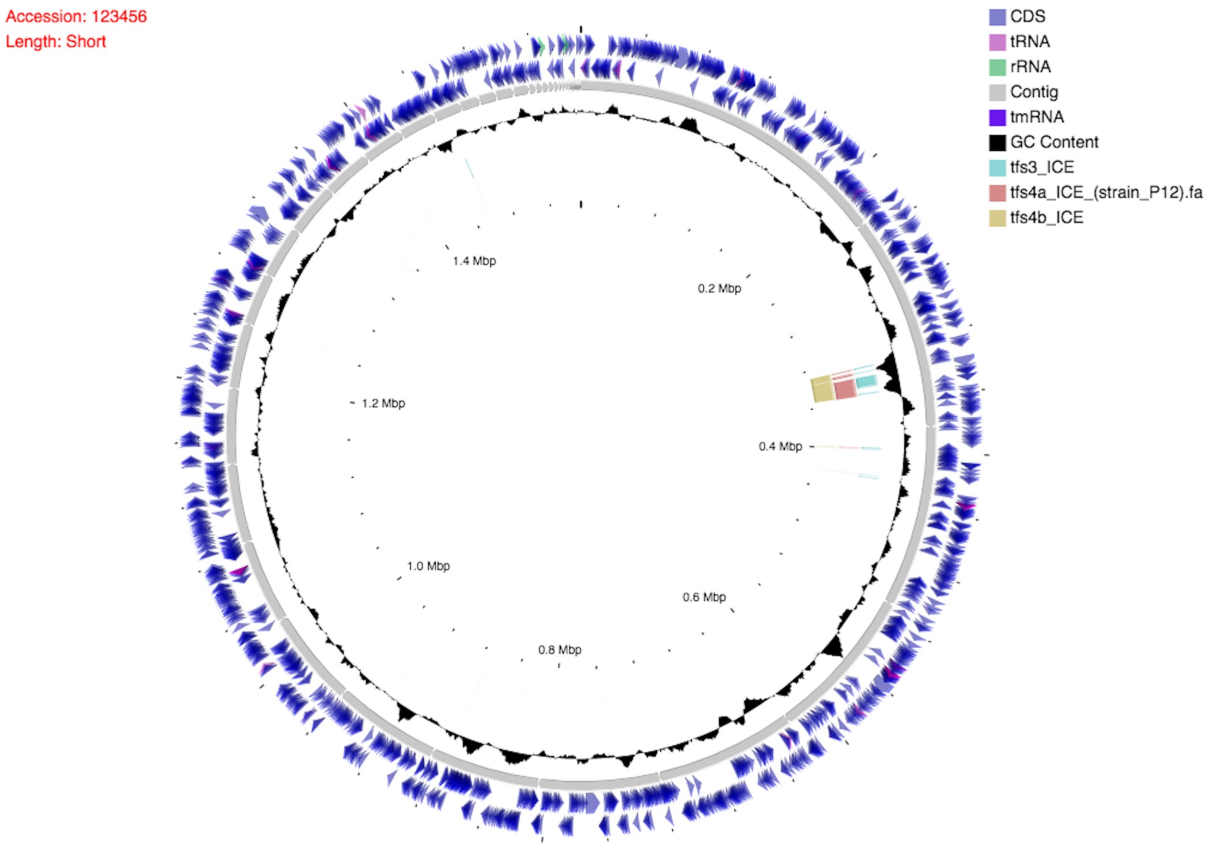
Supplementary Figure S6. *cag*PAI empty sites arrangement



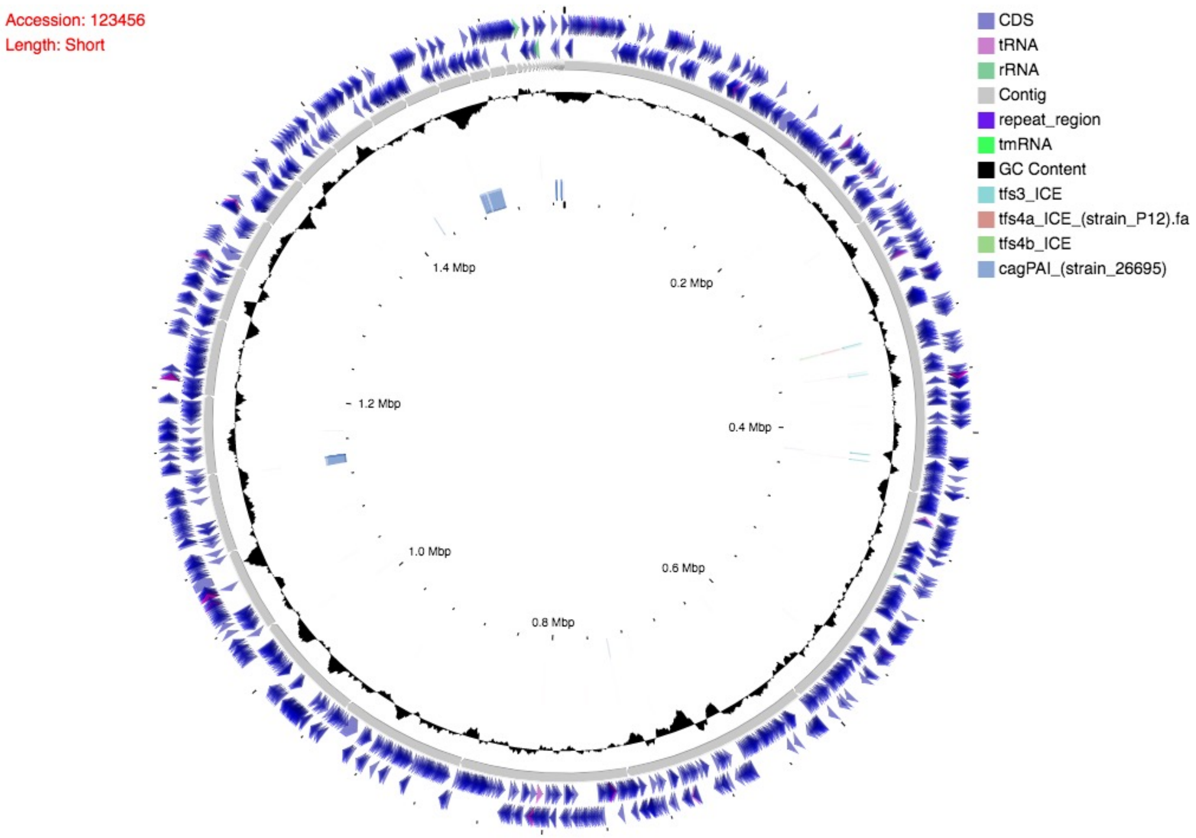
Supplementary Figure S7. The circular genome visualization of each strains. The type 4 secretion system shown as different color in legend and searched using BLASTN algorithm.



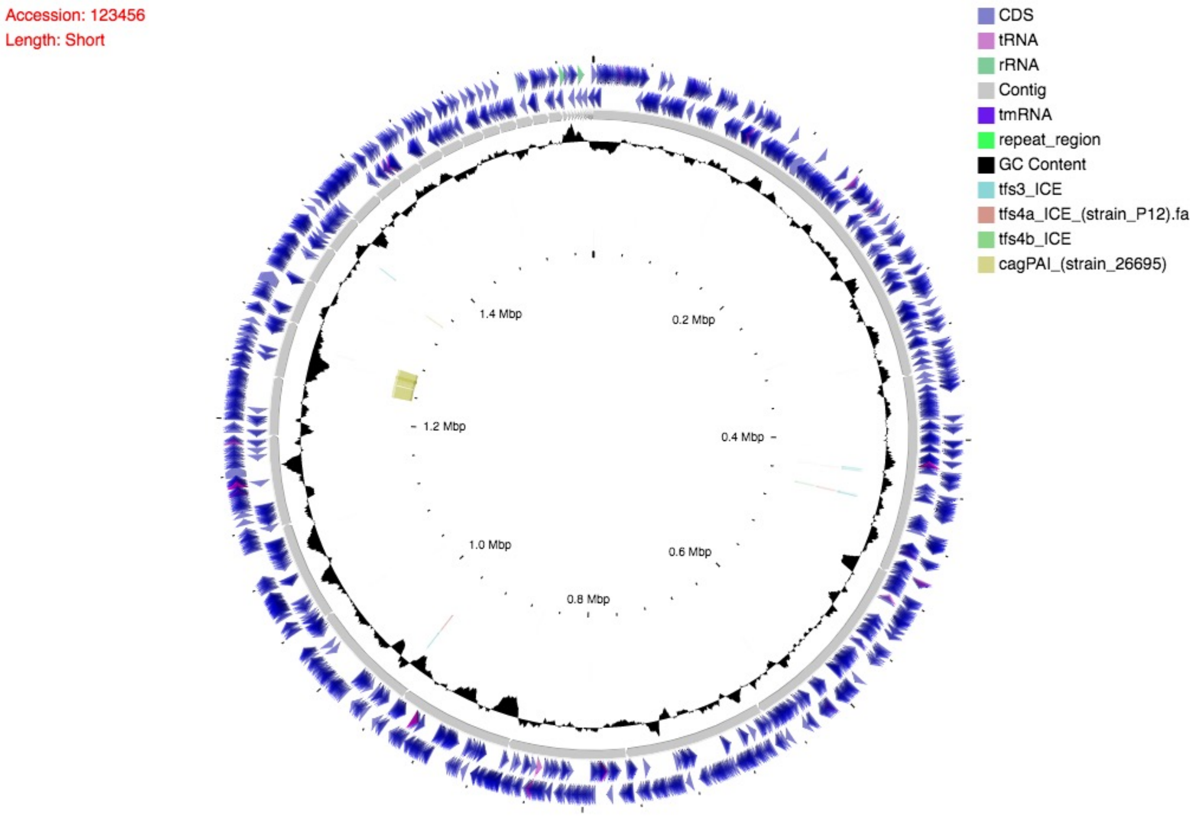
Helicobacter pylori SLK40



Helicobacter pylori SLK91

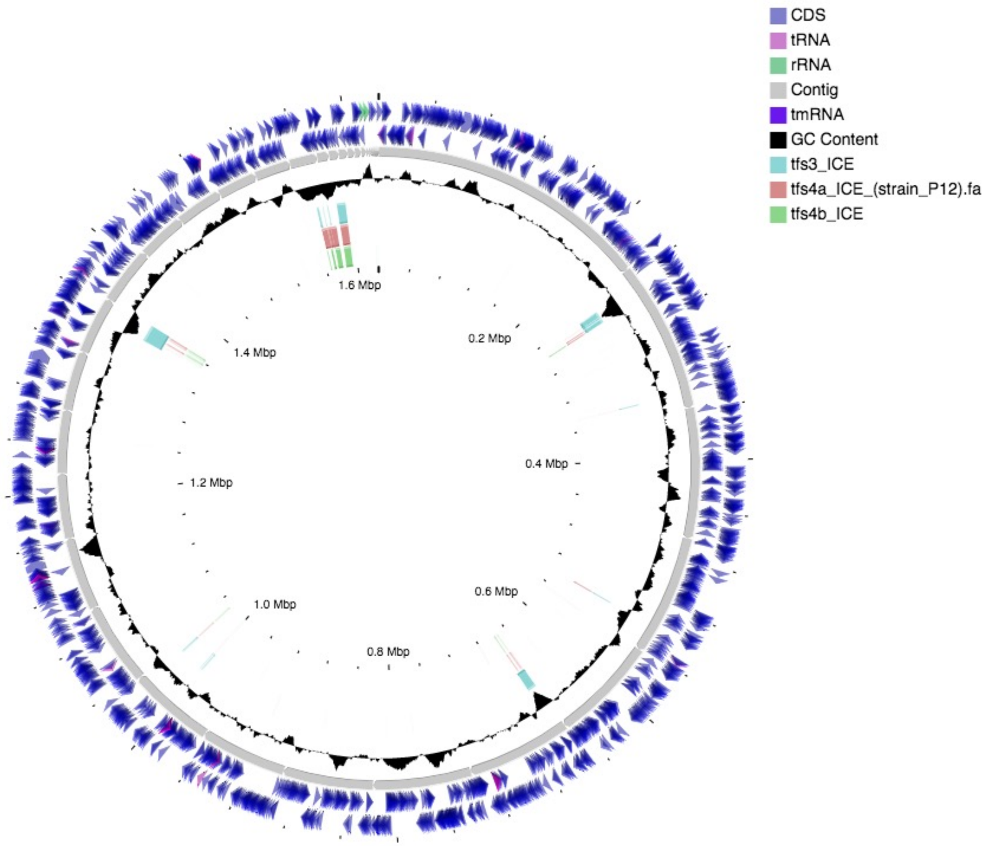


Helicobacter pylori SLK231



Helicobacter pylori SLK237

Accession: 123456
Length: Short



Helicobacter pylori SLK260