

Functions of Proteins

Sabrina Rasch

1. November 2021

Inhaltsverzeichnis

1	AGTU.c01	2
1.1	fadD	2
1.2	AGTU_00650	3
1.3	ropA_2	4
1.4	AGTU_01326	5
1.5	yjgN	6
2	AGTU.c02	7
2.1	secA	7
3	AGTU.p02	8
3.1	frmR	8
4	COTE.c01	9
4.1	COTE_04126	9
4.2	nasR	10
4.3	trkH	12
5	MISA.c01	13
5.1	ftsH_1	13
5.2	dedA_1	14
5.3	MISA_01549	15
5.4	puuP_3	16
5.5	srlR	17
5.6	srlR	17
6	OCAN.c02	18
6.1	yjiA	18

1 AGTU.c01

1.1 fadD

- Protein
 - Name: Long-chain-fatty-acid-CoA ligase
 - Uniprot ID: P69451
 - Start: 677068
 - End: 677808
 - DNA sequence: 1710
 - AA sequence: 570
- SNP
 - Microcosme.Replicate: 1.2
 - Position: 83654
 - Alternative: T
 - Reference: C
 - gen.fasta Reference: T
 - DNA Place: 887
 - AA Reference: L
 - AA Place: 296
- Uniprot ID
 - ID: P69451
 - Organism: Escherichia coli (strain K12)
 - Name: Long-chain-fatty-acid-CoA ligase
 - AA sequence: 561
 - Gene: fadD (oldD)
 - Predicted: 5/5
- BLAST DNA – / AA – Sequence
 - ID: A9CKB2
 - Organism: Agrobacterium fabrum (strain C58/ATCC 33970)
 - Name: Long-chain-fatty-acid-CoA ligase
 - AA sequence: 570
 - Gene: fadD
 - Predicted: 1/5
 - Agreement: 92.8%

1.2 AGTU_00650

- Protein
 - Name: Pentapeptide repeat protein Rfr32
 - Uniprot ID: B1WVN5
 - Start: 677068
 - End: 677808
 - DNA sequence: 741
 - AA sequence: 247
- SNP
 - Microcosme.Replicate: 3.2
 - Position: 677536
 - Alternative: C
 - Reference: G
 - gen.fasta Reference: G
 - DNA Place: 469
 - AA Reference: A
 - AA Place: 157
- Uniprot ID
 - ID: B1WVN5
 - Organism: *Crocospaera subtropica* (strain ATCC 51142 / BH68) (*Cyanosphaera* sp. (strain ATCC 51142))
 - Name: Pentapeptide repeat protein Rfr32
 - AA sequence: 179
 - Gene: rfr32
 - Predicted: 3/5
- BLAST DNA – / AA – Sequence
 - ID: Q7D066
 - Organism: *Agrobacterium fabrum* (strain C58 / ATCC 33970) (*Agrobacterium tumefaciens* (strain C58))
 - Name: Uncharacterized protein
 - AA sequence: 247
 - Gene:
 - Predicted: 1/5
 - Agreement: 93.9%

1.3 ropA_2

- Protein
 - Name: Outer membrane protein IIIA
 - Uniprot ID: Q05811
 - Start: 703145
 - End: 704191
 - DNA sequence: 1047
 - AA sequence: 349
- Uniprot ID
 - ID: Q05811
 - Organism: *Rhizobium leguminosarum* bv. *viciae*
 - Name: Outer membrane protein IIIA
 - AA sequence: 366
 - Gene: ropA
 - Predicted: 3/5
- SNP
 - Microcosme.Replicate: 1.1
 - Position: 703340
 - Alternative: T
 - Reference: C
 - gen.fasta Reference: C
 - DNA Place: 196
 - AA Reference: Q
 - AA Place: 66
- BLAST DNA – / AA – Sequence
 - ID: A0A1R3U7T0
 - Organism: *Agrobacterium rosae*
 - Name: Porin
 - AA sequence: 346
 - Gene:
 - Predicted: 2/5
 - Agreement: 74%

1.4 AGTU_01326

- Protein
 - Name: hypothetical protein
 - Uniprot ID: NA
 - Start: 1372185
 - End: 1372541
 - DNA sequence: 357
 - AA sequence: 119
- Uniprot ID
 - ID:
 - Organism:
 - Name:
 - AA sequence:
 - Gene:
 - Predicted:
- SNP
 - Microcosme.Replicate: 4.5
 - Position: 1372439
 - Alternative: A
 - Reference: C
 - gen.fasta Reference: C
 - DNA Place: 255
 - AA Reference: S
 - AA Place: 85
- BLAST DNA – / AA – Sequence
 - ID: A0A2L0W7V9
 - Organism: *Rhizobium* sp. NXC24
 - Name: Uncharacterised protein
 - AA sequence: 125
 - Gene:
 - Predicted: 1/5
 - Agreement: 63.1%

1.5 yjgN

- Protein
 - Name: Inner membrane protein YjgN
 - Uniprot ID: P39338
 - Start: 2863278
 - End: 2864348
 - DNA sequence: 1071
 - AA sequence: 357
- Uniprot ID
 - ID: P39338
 - Organism: *Escherichia coli* (strain K12)
 - Name: Inner membrane protein YjgN
 - AA sequence: 398
 - Gene: yjgN
 - Predicted: 2/5
- SNP
 - Microcosme.Replicate: 3.4
 - Position: 2863371
 - Alternative: T
 - Reference: G
 - gen.fasta Reference: A
 - DNA Place: 94
 - AA Reference: N
 - AA Place: 32
- BLAST DNA – / AA – Sequence
 - ID: A9CKK5
 - Organism: *Agrobacterium fabrum* (strain C58 / ATCC 33970) (*Agrobacterium tumefaciens* (strain C58))
 - Name: Uncharacterised protein
 - AA sequence: 367
 - Gene:
 - Predicted: 1/5
 - Agreement: 76.6%

2 AGTU.c02

2.1 secA

- Protein
 - Name: Protein translocase subunit SecA
 - Uniprot ID: P52966
 - Start: 553939
 - End: 556656
 - DNA sequence: 2718
 - AA sequence: 906
- Uniprot ID
 - ID: P52966
 - Organism: Rhodobacter capsulatus (Rhodopseudomonas capsulata)
 - Name: Protein translocase subunit SecA
 - AA sequence: 904
 - Gene: secA
 - Predicted: 4/5
- SNP
 - Microcosme.Replicate: 4.1
 - Position: 555455
 - Alternative: T
 - Reference: G
 - gen.fasta Reference: G
 - DNA Place: 1517
 - AA Reference: G
 - AA Place: 506
- BLAST DNA – / AA – Sequence
 - ID: Q7CSN9
 - Organism: Agrobacterium fabrum (strain C58 / ATCC 33970) (Agrobacterium tumefaciens (strain C58))
 - Name: Protein translocase subunit SecA
 - AA sequence: 902
 - Gene: secA
 - Predicted: 3/5
 - Agreement: 97.2%

3 AGTU.p02

3.1 frmR

- Protein
 - Name: Transcriptional repressor FrmR
 - Uniprot ID: P0AAP3
 - Start: 181281
 - End: 181562
 - DNA sequence: 282
 - AA sequence: 94
- SNP
 - Microcosme.Replicate: 4.4
 - Position: 181352
 - Alternative: A
 - Reference: G
 - gen.fasta Reference: T
 - DNA Place: 72
 - AA Reference: A
 - AA Place: 24
- Uniprot ID
 - ID: P0AAP3
 - Organism: Escherichia coli (strain K12)
 - Name: Transcriptional repressor Frm
 - AA sequence: 91
 - Gene: frmR
 - Predicted: 3/5
- BLAST DNA – / AA – Sequence
 - ID: A0A4Q2XBI5
 - Organism: Spirochaetia bacterium
 - Name: Metal/formaldehyde-sensitive transcriptional repressor
 - AA sequence: 93
 - Gene:
 - Predicted: 1/5
 - Agreement: 83.9%

4 COTE.c01

4.1 COTE_04126

- Protein
 - Name: hypothetical protein
 - Uniprot ID: NA
 - Start: 4436374
 - End: 4436670
 - DNA sequence: 297
 - AA sequence: 99
- Uniprot ID
 - ID:
 - Organism:
 - Name:
 - AA sequence:
 - Gene:
 - Predicted:
- SNP
 - Microcosme.Replicate: 2.5
 - Position: 4436504
 - Alternative: G
 - Reference: T
 - gen.fasta Reference: A
 - DNA Place: 131
 - AA Reference: K
 - AA Place: 44
- BLAST DNA – / AA – Sequence
 - ID: D0IZ09
 - Organism: Comamonas testosteroni (strain CNB-2)
 - Name: Uncharacterised protein
 - AA sequence: 153
 - Gene:
 - Predicted: 1/5
 - Agreement: 81.5%

4.2 nasR

- Protein
 - Name: Nitrate regulatory protein
 - Uniprot ID: Q48468
 - Start: 5465708
 - End: 5466982
 - DNA sequence: 1275
 - AA sequence: 425
- SNP
 - Microcosme.Replicate: 2.1
 - Position: 5466695
 - Alternative: C
 - Reference: A
 - gen.fasta Reference: G
 - DNA Place: 988
 - AA Reference: G
 - AA Place: 330
- SNP
 - Microcosme.Replicate: 2.2, 2.3
 - Position: 5466897
 - Alternative: C
 - Reference: A
 - gen.fasta Reference: C
 - DNA Place: 1190
 - AA Reference: A
 - AA Place: 397
- SNP
 - Microcosme.Replicate: 2.5
 - Position: 5466537
 - Alternative: G
 - Reference: T
 - gen.fasta Reference: A
 - DNA Place: 830
 - AA Reference: E
 - AA Place: 277
- SNP
 - Microcosme.Replicate: 3.3
 - Position: 5466905
 - Alternative: T
 - Reference: G
 - gen.fasta Reference: C
 - DNA Place: 1198
 - AA Reference: Q
 - AA Place: 400
- SNP
 - Microcosme.Replicate: 3.4
 - Position: 5466552
 - Alternative: A
 - Reference: G
 - gen.fasta Reference: G
 - DNA Place: 845
 - AA Reference: C
 - AA Place: 282

- Uniprot ID
 - ID: Q48468
 - Organism: *Klebsiella oxytoca*
 - Name: Nitrate regulatory protein
 - AA sequence: 396
 - Gene: nasR
 - Predicted: 2/5
- BLAST DNA – / AA – Sequence
 - ID: D0IZ81
 - Organism: *Comamonas testosteroni*
(strain CNB-2)
 - Name: Response regulator receiver
and ANTAR domain protein
 - AA sequence: 431
 - Gene:
 - Predicted: 1/5
 - Agreement: 96.7%

4.3 trkH

- Protein
 - Name: Trk system potassium uptake protein TrkH
 - Uniprot ID: E1V6C5
 - Start: 5643413
 - End: 5644879
 - DNA sequence: 1467
 - AA sequence: 489
- SNP
 - Microcosme.Replicate: 2.3
 - Position: 5643748
 - Alternative: C
 - Reference: G
 - gen.fasta Reference: C
 - DNA Place: 336
 - AA Reference: G
 - AA Place: 112
- Uniprot ID
 - ID: E1V6C5
 - Organism: Halomonas elongata (strain ATCC 33173 / DSM 2581 / NBRC 15536 / NCIMB 2198 / 1H9)
 - Name: Trk system potassium uptake protein TrkH
 - AA sequence: 482
 - Gene: trkH
 - Predicted: 2/5
- BLAST DNA – / AA – Sequence
 - ID: D0IW70
 - Organism: Comamonas testosteroni (strain CNB-2)
 - Name: Trk system potassium uptake protein
 - AA sequence: 454
 - Gene:
 - Predicted: 2/5
 - Agreement: 97.5%

5 MISA.c01

5.1 ftsH_1

- Protein
 - Name: ATP-dependent zinc metallo-protease FtsH
 - Uniprot ID: P9WQN3
 - Start: 697714
 - End: 699714
 - DNA sequence: 2001
 - AA sequence: 667
- SNP
 - Microcosme.Replicate: 4.3
 - Position: 699272
 - Alternative: T
 - Reference: C
 - gen.fasta Reference: C
 - DNA Place: 1559
 - AA Reference: T
 - AA Place: 520
- Uniprot ID
 - ID: P9WQN3
 - Organism: *Mycobacterium tuberculosis* (strain ATCC 25618 / H37Rv)
 - Name: ATP-dependent zinc metallo-protease FtsH
 - AA sequence: 760
 - Gene: ftsH
 - Predicted: 5/5
- BLAST DNA – / AA – Sequence
 - ID: T5KSX8
 - Organism: *Microbacterium mageritense* MF109
 - Name: ATP-dependent zinc metallo-protease FtsH
 - AA sequence: 666
 - Gene: ftsH
 - Predicted: 3/5
 - Agreement: 97%

5.2 dedA_1

- Protein
 - Name: Protein DedA
 - Uniprot ID: P0ABP6
 - Start: 1281677
 - End: 1282456
 - DNA sequence: 780
 - AA sequence: 260
- Uniprot ID
 - ID: P0ABP6
 - Organism: *Escherichia coli* (strain K12)
 - Name: Protein DedA
 - AA sequence: 219
 - Gene: dedA
 - Predicted: 2/5
- SNP
 - Microcosme.Replicate: 4.5
 - Position: 1281972
 - Alternative: G
 - Reference: T
 - gen.fasta Reference: G
 - DNA Place: 296
 - AA Reference: G
 - AA Place: 99
- BLAST DNA – / AA – Sequence
 - ID: T5KJN8
 - Organism: *Microbacterium maritypi* cum MF109
 - Name: Membrane protein
 - AA sequence: 259
 - Gene:
 - Predicted: 1/5
 - Agreement: 98.5%

5.3 MISA_01549

- Protein
 - Name: PhoH-like protein
 - Uniprot ID: P9WIA3
 - Start: 1628109
 - End: 1629221
 - DNA sequence: 1113
 - AA sequence: 371
- Uniprot ID
 - ID: P9WIA3
 - Organism: *Mycobacterium tuberculosis* H37Rv (strain ATCC 25618 / H37Rv)
 - Name: PhoH-like protein
 - AA sequence: 352
 - Gene:
 - Predicted: 2/5
- SNP
 - Microcosme.Replicate: 3.3
 - Position: 1628511
 - Alternative: G
 - Reference: T
 - gen.fasta Reference: T
 - DNA Place: 403
 - AA Reference: Y
 - AA Place: 135
- BLAST DNA – / AA – Sequence
 - ID: T5KJR3
 - Organism: *Microbacterium maritipiscum* MF109
 - Name: Phosphate starvation protein PhoH
 - AA sequence: 370
 - Gene:
 - Predicted: 1/5
 - Agreement: 97%

5.4 puuP_3

- Protein
 - Name: Putrescine importer PuuP
 - Uniprot ID: P76037
 - Start: 2587129
 - End: 2588556
 - DNA sequence: 1428
 - AA sequence: 476
- Uniprot ID
 - ID: P76037
 - Organism: *Escherichia coli* (strain K12)
 - Name: Putrescine importer PuuP
 - AA sequence: 461
 - Gene: puuP (ycjJ)
 - Predicted: 5/5
- SNP
 - Microcosme.Replicate: 3.5
 - Position: 2587322
 - Alternative: C
 - Reference: A
 - gen.fasta Reference: T
 - DNA Place: 194
 - AA Reference: F
 - AA Place: 65
- BLAST DNA – / AA – Sequence
 - ID: A0A543BPR9
 - Organism: *Microbacterium saperdae*
 - Name: Putrescine:proton symporter (AAT family)
 - AA sequence: 475
 - Gene:
 - Predicted: 1/5
 - Agreement: 96.6%

5.5 srlR

5.6 srlR

- Protein
 - Name: Glucitol operon repressor
 - Uniprot ID: P15082
 - Start: 3643263
 - End: 3644012
 - DNA sequence: 750
 - AA sequence: 250
- SNP
 - Microcosme.Replicate: 4.5
 - Position: 3643534
 - Alternative: C
 - Reference: A
 - gen.fasta Reference: A
 - DNA Place: 272
 - AA Reference: D
 - AA Place: 91
- Uniprot ID
 - ID: P15082
 - Organism: Escherichia coli (strain K12)
 - Name: Glucitol operon repressor
 - AA sequence: 257
 - Gene: srlR (gutR)
 - Predicted: 2/5
- BLAST DNA – / AA – Sequence
 - ID: T5KNG3
 - Organism: Microbacterium maritypicum MF109
 - Name: HTH deoR-type domain-containing protein
 - AA sequence: 249
 - Gene:
 - Predicted: 1/5
 - Agreement: 79.9%

6 OCAN.c02

6.1 yjiA

- Protein
 - Name: putative GTP-binding protein YjiA
 - Uniprot ID: P24203
 - Start: 2115807
 - End: 2116997
 - DNA sequence: 1191
 - AA sequence: 397
- SNP
 - Microcosme.Replicate: 4.1, 4.3, 4.4, 4.5
 - Position: 2116572
 - Alternative: C
 - Reference: G
 - gen.fasta Reference: G
 - DNA Place: 766
 - AA Reference: D
 - AA Place: 256
- Uniprot ID
 - ID: P24203
 - Organism: *Escherichia coli* (strain K12)
 - Name: P-loop guanosine triphosphatase YjiA
 - AA sequence: 318
 - Gene: yjiA
 - Predicted: 5/5
- BLAST DNA – / AA – Sequence
 - ID: A6WXC0
 - Organism: *Ochrobactrum anthropi* (strain ATCC 49188 / DSM 6882 / JCM 21032 / NBRC 15819 / NCTC 12168)
 - Name: Cobalamin synthesis protein P47K
 - AA sequence: 382
 - Gene:
 - Predicted: 1/5
 - Agreement: 94%