

## Novel short ORFs detected.

### Differentially expressed novel peptides

The six differentially expressed ORFs (**D1-D6**) are listed first with nucleotide sequence, amino acid sequence and JPRED structure prediction. Nucleotide sequence includes 10 upstream bases to show possible Shine-Dalgarno sequence. Start codon is shown in **bold**. The Blast search information is described, and the top hit is listed.

Fasta header : >Position of start of ORF|Position of end of ORF|Strand|Fold change (log<sub>2</sub>)|Regulation (up/down)

### Differentially expressed small peptides:

#### D1

BLAST : no obvious similarities

NT seq

>2009278|2009359|-|1.43|diff\_non|Upregulated  
GGGATTTGCTAT**G**AAAGATCGCAGCGAAAATATATTGTTACGAATCCCGC  
TGATTATGATAAATATCCGATATGGTTAAAAAAGTATGTAA

AA seq

>2009278|2009359|-|1.43|diff\_non  
MKDRSENILLRIPLIMINIPIWLKSM

Structure:

Query: MKDRSENILLRIPLIMINIPIWLKSM  
Jpred: -----H-HHHHHHHHHHHHHHHHHHH--  
Conf: 99986116888899999999987289

#### D2

BLAST : high similarity to ref|ZP\_04872718.1| conserved hypothetical protein [Escherichia sp. 1\_1\_43]

NT seq

>3634009|3634096|-|1.41|prime3|Upregulated  
GGATGGAAAGGT**G**ATTGAAACTCACTCAGTGGCCAGATCTTTATGAAAT  
ATGAACAGGTTGAAAAAACAGTAACTTTGCTGTTTTTATACTTAA

AA seq

>3634009|3634096|-|1.41|prime3  
VIENSLSGQIFMKYEQVEKNSNFAVFYT

Structure :

Query: VIENSLSGQIFMKYEQVEKNSNFAVFYT  
Jpred: -----EEEEEE-----EEEE--  
Conf: 6888888607877760478873688736

### D3

BLAST : high similarity in the N-terminal end to ref[ZP\_06167183.1] glyoxalase I, nickel isomerase [Klebsiella variicola At-22]

NT seq

>4529978|4530035|-|1.4|diff\_non|Downregulated

AGGAGATCCGATGTTTTCTTACATTATGCTGGGAACCAACGACCTGCCAC  
GCGCCATTACATTCTAA

AA seq

>4529978|4530035|-|1.4|diff\_non

MFSYIMLGTNDLPRAITF

No structure predicted - too short

### D4

BLAST : high similarity to ref[YP\_002394396.1] putative frameshift suppressor; KpLE2 phage-like element [Escherichia coli S88]

NT seq

>4533067|4533673|+|0.8|diff\_non|Downregulated

TGGTAGAGGCATGAAGCCAAATGGAGGGAGCCGTCATAACCCTCTTGAA  
GCTGAAACAATAGCAGCATGGCTGGTGGCACATAAGGATGATATAGAGC  
GCCATTACGGGGAACCACTGTACAAAGTTGTTGGTGTGTGACGCCATTTT  
CGGCGCAGGTTAATGCTATCAAATGTCATTGCGTAAGCTGGAAATTAAT  
GGCAAAGACGAGCAGGGTTTACTGACAGTGGGGACTGTGCACTCCTTGCA  
GGGGGCGGAAAGGGCGATTGTTCTCTTTTCTCCTGTCTACTCAAACATGA  
AGATGGCAGGTTTCTTGATAGCAACAGCACTATCCTCAATGTTGCTGTCTC  
ACGCGCTAAAGATAGTTTTCTGGTATTTGGCGATATGGACCTTATCGAGAT  
GCAGCCCGCATTTTCGCCGCGAGGGTTATTGGCAAATATCTCTTTTCTTC  
AGATAATAATGCGTTGCAATTTGAGTTTCAGAAACGACAGGATTTAATTT  
CCGCACATACAAAATTTCAACGCTGCATGGTGTGGAGCAGCATGATGAA  
TTTTTGAACAAGACTCTTGCCGGAGCTCAAAGAAAATAACGATTATTC  
CCCCGAGATATAG

AA seq

>4533067|4533673|+|0.8|diff\_non

MKPNGGSRHNPLEAETIAAWLV AHKDDIERHYGEPLYKVVGVVTPFSAQVN  
AIKMSLRKLEINGKDEQGLLTVGTVHSLQGAERAIVLFSPVYSKHEDGRFLDS  
NSTILNVAVSRAKDSFLVFGDMDLIEMQPAFSPRGLLAKYLFSSDNNALQFEF  
QKRQDLISAHTQISTLHGVEQHDEFLNKTLAGAQQKITIISPEI

Structure:

Query:

MKPNGGSRHNPLEAETIAAWLV AHKDDIERHYGEPLYKVVGVVTPFSAQVN  
AIKMSLRKL



Query:  
MEDYWLRLPIRFYESERRNFLGTLFSEGGLPFRLKESDSRFLAVFSRILGQYEQ  
AKQSGF  
Jpred: -----EEEE-----EEEEEEE-----HHHH-----HHHHHHHHHHHHHHHHHH-----  
Conf: 998557500062488873688864057765000004787557899999999862068887

Query:  
SALSLARAVIEKSALPTVFS EDTSVELISHMADNLNSLVLT HN LNHKEPVQQL  
EKVHPT  
Jpred: HHHHHHHHHHHH---HHHHH---HHHHHHHHHHHHHHHHHHHHHHHHHHHH-----  
HHHHHHHHH---  
Conf: 0688999885007501200777468999999999999999860578875078898874266

Query:  
WRSEFPIPLDDETGTHFLNGLLCAASVEAKPRLQKNKSTRCQFYWSEKHPDE  
LRVIVSLP  
Jpred: -----HHHHHHHHHHHHHHHHHHHHHHHHHHHH-----EEEEEEEE-----  
EEEEEEEE-  
Conf: 6556777876517999999999999998750678877725889987167604789999844

Query:  
DEVSPVTSEPSTTRFELAICEDGEEVSGLDLYRWRYDYTKRRNCNAK  
Jpred: -EEEE-----EEEEEEE-----HHH-----  
Conf: 308880268877750578775046005776100000500544567889

## D6

BLAST : high similarity to ref|YP\_002415529.1| hypothetical protein ECUMN\_4952  
[Escherichia coli UMN026]

### NT seq

>4571523|4571883|+|1.17|diff\_non|Downregulated

GCTGGAGGGATGATTGTTGGGAGCATTAACTTGATGACAGTGAAATTGC  
CATTGGTGAGGTGCCGTTAACCTTCATCGTTGATGCAGATCAATGGCTGTT  
ACAGGGACAGGCTTCTTGCAGTGTGCGAAGCAGCGATGTTCTGATTGTGC  
TCCCTCGGGATAATAGCAATGTTGCTGGTTTTGATGGCCAATCGAGGGCA  
GTAAACGTATTAGGACTAAAAGCACTACCTGTGAAGGGATGTCAGGACGT  
TACAGTTACAGCAAATGAACTTATCGGATTCGCACCGGAAGAGAACAAA  
TCAGCATCGGAAGGTTTGCTCTAAATGGAAAGCGTGCAAGCTGGGTTTGT  
CATCCAGATGAAACATAG

### AA seq

>4571523|4571883|+|1.17|diff\_non

MIVGSIKLLDSEIAIGEVPLTFIVDADQWLLQGQASCSVRSSDVLIVLPRDNSN  
VAGFDG  
QSRAVNVLGLKALPVKGCQDVTVTANETYRIRRTGREQISIGRFALNGKRASW  
VCHPDET

Structure:

Query:

MIVGSIKLDDSEIAIGEVPLTFIVDADQWLLQGQASCSVRSSDVLIVLPRDNSN  
VAGFDG

Jpred: -EEEE-----EEEEEE--EEEEEE-----E---EEEEEE-----

Conf: 62652004776545776200898507515677347750237606888536777777777

Query:

QSRVNVLGLKALPVKGCQDVTVTANETYRIRTGREQISIGRFALNGKRASW  
VCHPDET

Jpred: ----EEEE-----EEE---EEEEEE-----EE--EEE--EEEEEE-----

Conf: 76650000256770077762577068558998547850500077705046656037899

## Similarly expressed small peptides

The 17 differentially similarly expressed ORFs (S1-S17) are listed with nucleotide sequence, amino acid sequence and JPRED structure prediction. The nucleotide sequence includes 10 nucleotides upstream of the start codon to show possible Shine-Dalgarno sequence, the start codon is in **bold**.

Fasta header: > Position of start of ORF|end position of ORF|strand

### S1

BLAST : high similarity to nucleotide sugar synthetase [Escherichia coli W3110]

NT seq

>2101098|2101158|-

TGATTACATTT**TG**TTTTTGAATCCCGATATCATCATGAAGCATGATGATT  
GCTGACATATATTAATAT

AA seq

>2101098|2101158|-

LFLNPDIIIMKHDDLLTYIKY

Structure :

Query: LFLNPDIIIMKHDDLLTYIKY

Jpred: -----EEE----HHHH---

Conf: 87467007703642322079

### S2

BLAST high similarity to gb|AAB40749.1|outer membrane porin protein [Escherichia coli]

NT seq

>575577|576048|-

AGGGCTCACGAT**TG**AAAAAATTAACAGTGGCAATTTCTGCTGTAGCTGCAT  
CAGTACTAATGGCGATGTCTGCTCAGGCAGCTGAAATTTATAATAAAGAC  
AGTAACAAGCTGGATCTGTACGGGAAAGTTAATGCTAAGCACTACTTCTC  
CTCTAATGATGCAGATGATGGTGATACTACTTATGCCCGTCTTGGCTTCAA  
AGGTGAAACCCAAATCAACGATCAACTGACTGGTTTCGGTCAGTGGGAAT  
ATGAATTCAAAGGCAACCGCGCTGAATCTCAAGGTTCCCAAAGATAAA  
ACCCGTCTTGCCTTCGCTGGCCTGAAATTCGGTGACTACGGCTCCATCGAT  
TACGGCCGTAACACTACGGTGTAGCATAACGACATCGGTGCGTGGACTGACGT  
CCTGCCAGAATTCGGTGGTGACACTTGGACTCAAACCGACGTGTTTCATGA  
CTCAACGTGCAACTGGTGTGCAACCTAT

AA seq

>575577|576048|-

MKKLTVAISAVAASVLMAMSAQAAEIYNKDSNKLDLYGKVNAKHYFSSND  
ADDGDTTYARLGFKGETQINDQLTGFGQWEYEFKGNRAESQGSSKDKTRLA  
FAGLKFGDYGSIDYGRNYGVAYDIGAWTDVLPFEGGDTWTQTDVFMTQRAT  
GVATY

Structure :



Conf: 99999999999870565899999999999999860589

#### S4

BLAST: no obvious similarities

NT seq

>3358750|3358834|-

TAACAAGTAATTGAGACGGATTGTAATGAGAAATTGCTTGATGTGCTTC  
ACTCATCAGACTTGGTGAACAAAGCACAATTTATTGAAAATTAT

AA seq

>3358750|3358834|-

LRRIVNEKLLDVLHSSDLVNKAQFIENY

Structure :

Query: LRRIVNEKLLDVLHSSDLVNKAQFIENY

Jpred: --HHHHHHHHHHHH----HHHHHHHH--

Conf: 9848999999875378855774776279

#### S5

BLAST : high similarity to malate dehydrogenase [Escherichia coli UTI89]

NT seq

>3382305|3382356|-

AACGGCAATTTTGTGGATTAAGGTCGCGGCAGCGGAGCAACATATCTTAG  
TTTATCAATAT

AA seq

>3382305|3382356|-

LWIKVAAAEQHILVYQY

Structure : Too short to predict

#### S6

BLAST : high similarity to hypothetical protein EcE24377A\_0429 [Escherichia coli E24377A]

NT seq

>418433|418529|-

TGAAAAAGACATGGCTAATCATGCGAACAAAAAATCACCAATCGTTGTA  
AAATCTAACGCCCGTATTTTGACTGCAAAGAAAATAAGCCAGCATCGACT  
GGCTTAT

AA seq

>418433|418529|-

MANHANKKSPIVVKSNARILTAKKISQHRLAY

Structure :



Query: MANHANKKSPIVVKSNARILTAKKISQHRLAY

Jpred: -----EEEE----EEEE-HH-H-----

Conf: 99888877736885478457612330057999

### S7

BLAST : exactly similar to: hypothetical protein EcF11\_2204 [Escherichia coli F11].

NT seq

>271775|271883|+

GTGTTTTGACATGTATGCGAAACACCGGGGCAGCAGCGTCTATCACGCTT  
GCGTTGCTGACTTCTGCCAACTTGC GGCAAGCAAGGATAAAGAGTGCGAC  
GGGCAGCCTCCTCAGTAT

AA seq

>271775|271883|+

MYAKHRGSSVYHACVADFCQLAASKDKECDGQPPQY

Structure :

Query: MYAKHRGSSVYHACVADFCQLAASKDKECDGQPPQY

Jpred: -----HHHHHHHHHHHHHH-----

Conf: 9866777750089999999870267777777888

### S8

BLAST : no obvious similarities

NT seq

>1014684|1014786|-

GGCATGTCGGTTGTACAAAAAATGTGAAGAGGAACAAATTTTTCCGCACG  
TTAGAGACAAAAAACCGCCTCAGGCAACATACCGGAGACGGTTATATTT  
ACAAATCTTTAT

AA seq

>1014684|1014786|-

LYKKCEEEQIFPHVRDKKTASGNIPETVIFTNLY

Structure :

Query: LYKKCEEEQIFPHVRDKKTASGNIPETVIFTNLY

Jpred: -----EEEE----

Conf: 9988877777777777777777777777777515635899

### S9

BLAST : no obvious similarities

NT seq

>1801151|1801202|-

ATAGGCTCTGATGAAGGTGTTGGTATGTCGCTATCGATGTGCAATCTATG  
AGGAGACTTAT

AA

>1801151|1801202|-

MKVLVCRYRCAIYEETY

Structure : Too short to predict

### **S10**

BLAST : no obvious similarities

NT seq

>1677411|1677483|-

TTACGAAGAGATG**T**GTGTCTGCATGCAATGGTCGCGGGGAAATTTGTGACTA  
AAAATGCGAAAGGCGGCATCAAGCCGCCTTAT

AA seq

>1677411|1677483|-

MCLHAMVAGKFVTKNAKGGIKPPY

Structure :

Query: MCLHAMVAGKFVTKNAKGGIKPPY

Jpred: ---HHHHH--EEE-----

Conf: 997466534536437777778899

### **S11**

BLAST : no obvious similarities

NT seq

>3672497|3672548|-

TCCCAAATAGATG**G**GCTGAAGAGGCTCTGGGGACGAGTTTCAAGAGGTTAT  
GTGCGCATTAT

AA seq

>3672497|3672548|-

MAEEALGTSFKRLCAHY

Structure: too short to predict

### **S12**

BLAST : no obvious similarities

NT seq

>892871|892937|-

TGCAATAAACG**T**G**C**ATATTTTTTATCGGCCGCGGTGTAAGAAAAGAGAACT  
TTCTCCCGCCAGCACTGCTCGCTAT

AA seq

>892871|892937|-

VHIFYRRGVRKENFLPASTARY

Structure :

Query: VHIFYRRGVRKENFLPASTARY

Jpred: ---H-----

Conf: 971014677777777778899

### S13

BLAST : exactly similar to hypothetical protein ECO26\_4577 [Escherichia coli O26:H11 str. 11368]

NT seq

>3632079|3632250|-

TAGTCACCAGATGGAGAAAGTATGGTTTTATAATAGTGAACCTTCTATAG  
ATATTC AAGAATATAATAGATTCCGCCGACAATACTACTGATACGTTTATCT  
TTACAATTATACCAGACAATAACCATGTGTTAAAATTATCGTCTCCCATTA  
CCGTTACTGTTGAGTGTAAGGGCGGTTAT

AA seq

>3632079|3632250|-

MEKVWFYNSEPSIDIQEYNRFADNTTDTFIFTIIPDNNHVLKLSSPITVTVECKG  
GY

Structure :

Query:

MEKVWFYNSEPSIDIQEYNRFADNTTDTFIFTIIPDNNHVLKLSSPITVTVECKG  
GY

Jpred: --EEEE-----HH-----EEEEEE---EEEE---EEEEEE---

Conf: 980440688765222000567777787058898548862788548508999844899

### S14

BLAST : similar to hypothetical protein CKO\_01089 [Citrobacter koseri ATCC BAA-895].

NT seq

>1956449|1956515|-

GTCGCCCAATATGATGCTTGCTCGTACCAGGCCCTGCAATTTCAACAGG  
GGCCTTTTTTATCCCTGAACAGTAT

AA seq

>1956449|1956515|-

MMLARTRPLQFQQGPFPIPEQY

Structure :

Query: MMLARTRPLQFQQGPFPIPEQY

Jpred: -----

Conf: 998887777777764257899

### S15

BLAST : exactly similar to hypothetical protein ECIAI1\_2752 [Escherichia coli IAI1].

NT seq

>2783625|2783718|  
TCTATATTTTCATGACTAACGCCATTTCAAGGAAGAAGTGCGGTGGGTGGT  
TTCGAATGAGGGCTGGTTCTATATTTCTGTTTTGGTACGTGGCACGCCACTG  
CAATGTATTTAT

AA seq  
>2783625|2783718|  
MTNAISRKKCGGWFRMRAGVLRGTPLQCIY

Structure :  
Query: MTNAISRKKCGGWFRMRAGVLRGTPLQCIY  
Jpred: ---HHHHHH---EEE--EEEE-----  
Conf: 9855777504670450010898036544126

**S16**  
BLAST : exactly similar to hypothetical protein ECIAI1\_2985 [Escherichia coli IAI1]

NT seq  
>2997966|2998023|  
CCTTTGCCAAATGAAATTAACCTCTTTCTCTATTACCACGTTTTTCCAGAA  
GCAAGAGATTGGGTAT

AA seq  
>2997966|2998023|  
MKLTSFSITTFQKQEIGY

Structure : Too short to predict

**S 17**  
BLAST : exactly similar to hypothetical protein E2348C\_1940 [Escherichia coli O127:H6 str. E2348/69].

NT seq  
>1899629|1899806|  
CGTCTCTGATATGTTTAACTCCCGTTTAAACAACCATGGAGTATAGGGCAGT  
AGCCCCGAGTATGGATCGTCACCGACGTCATTCAGCATCAGGCCTTTTAA  
CGCCTGCCTTTCTGGCACTCTATGCCGCACCTTTCGTTTGCATTTTGTTCGTT  
ACGCCTGCATTATTTCTGGCGTCGAATAGCTAT

AA seq  
>1899629|1899806|  
MFNSRLTTMEYRAVARSMRHRHFSIRPFNACLSGTLCRTRFRLHFVVTPALF  
LASNSY

Structure :  
Query:MFNSRLTTMEYRAVARSMRHRHFSIRPFNACLSGTLCRTRFRLHFVV  
TPALFLASNSY  
Jpred: ---EEEEHHHHHHHHHH-----H-HHH-HH---EEEEEE--HHH-----  
Conf: 85000000026889900000000232222000000000345178987320011057899