

Cluster AY - IrrE-like metallopeptidase

Genes related to MidnightRain_54

Challenge/issue:

- Glimmer/GeneMark predictions often call starts in the middle of the peptidase domain, calling into question whether the protein could function as a peptidase.
- Coding potential drops off, but does the ORF extend up to the upstream gene?

Auto-annotation and coding potential

Typically a gap in coding potential between this gene and its upstream gene

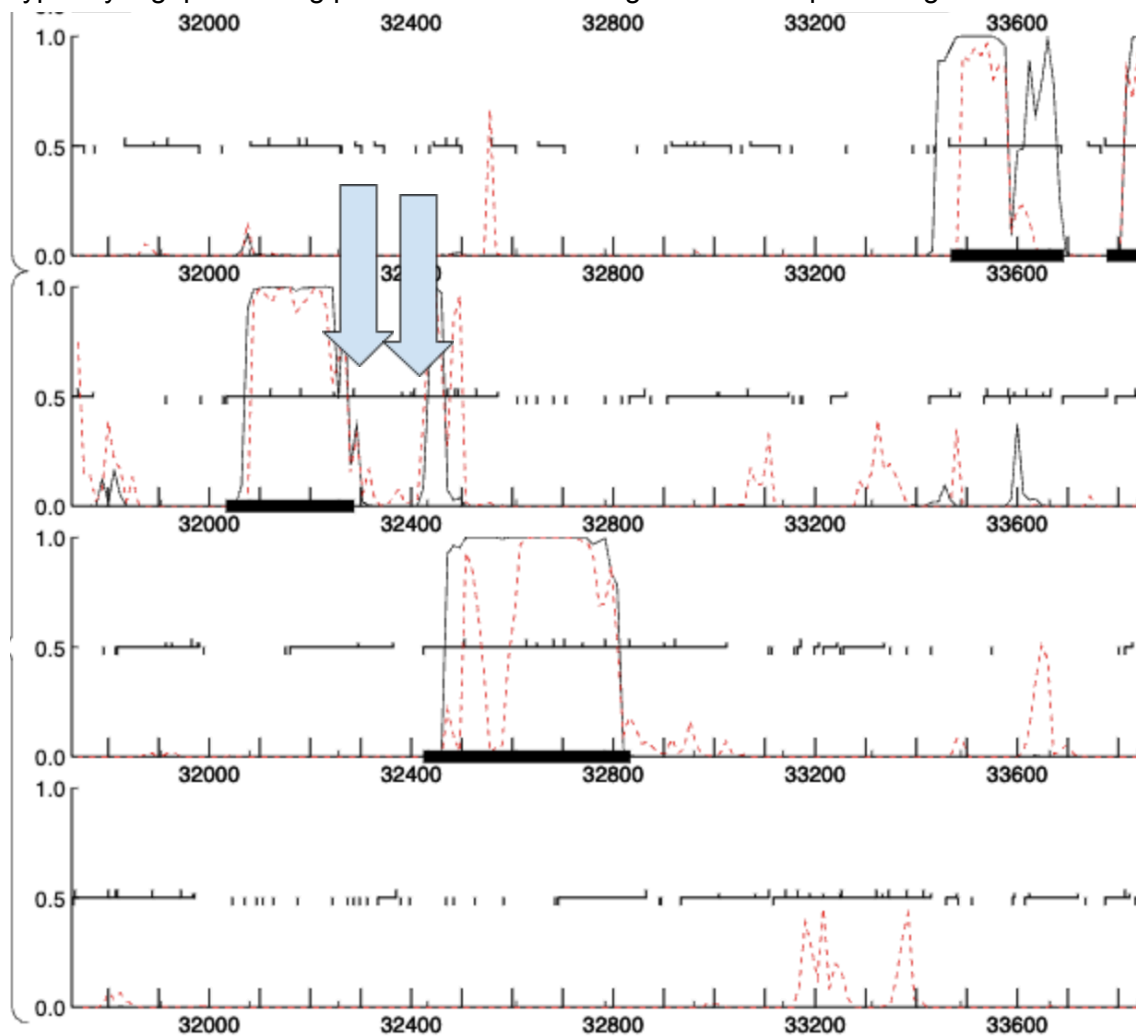


Figure above: Genemark Sashimi genes 52, 53, & 54.

We will call the autoannotated start the SHORT start, or short protein. (~32300 above)

We will call the start that closes the gap the LONG start, or the long protein. (~32400 above)

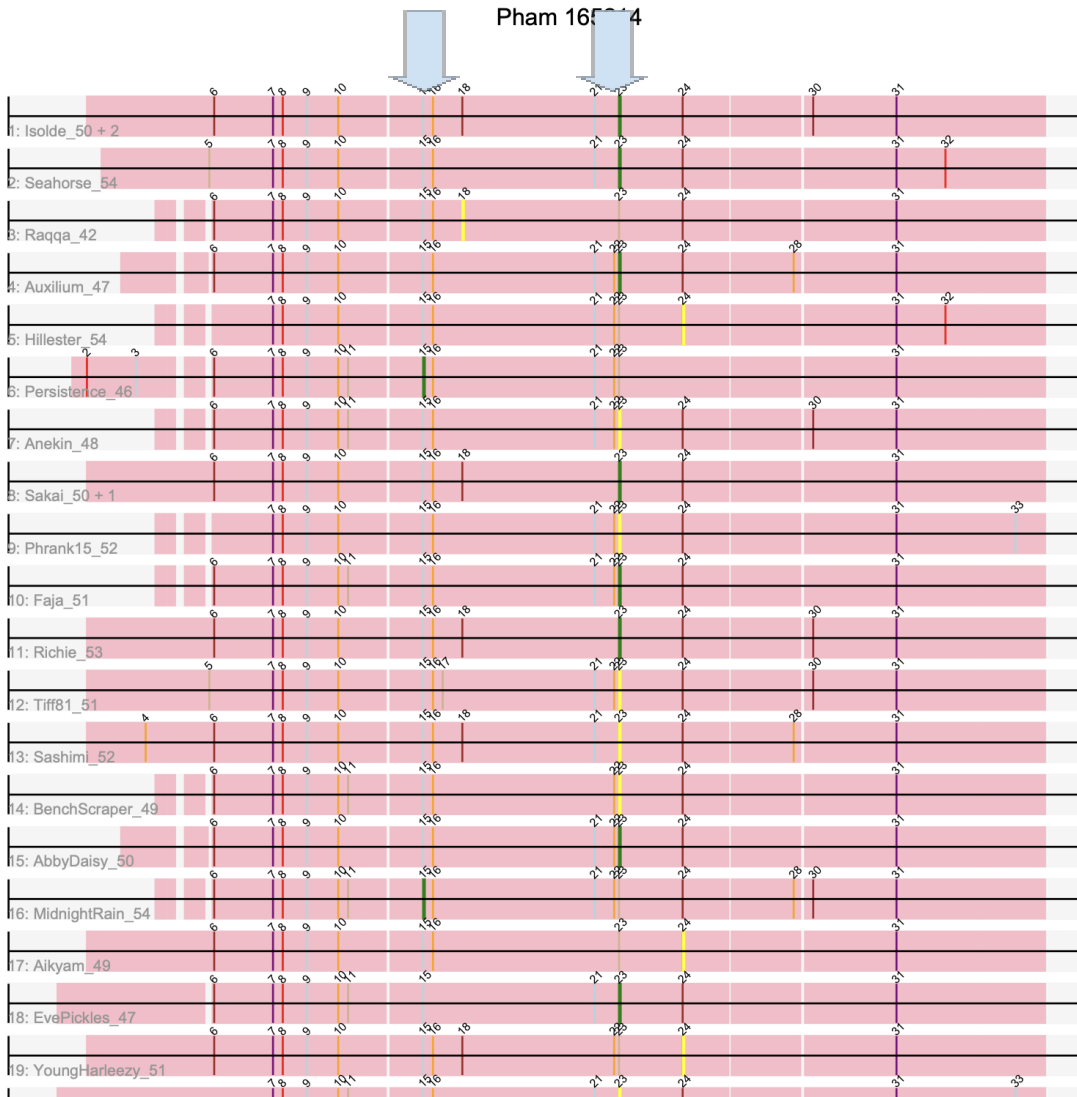
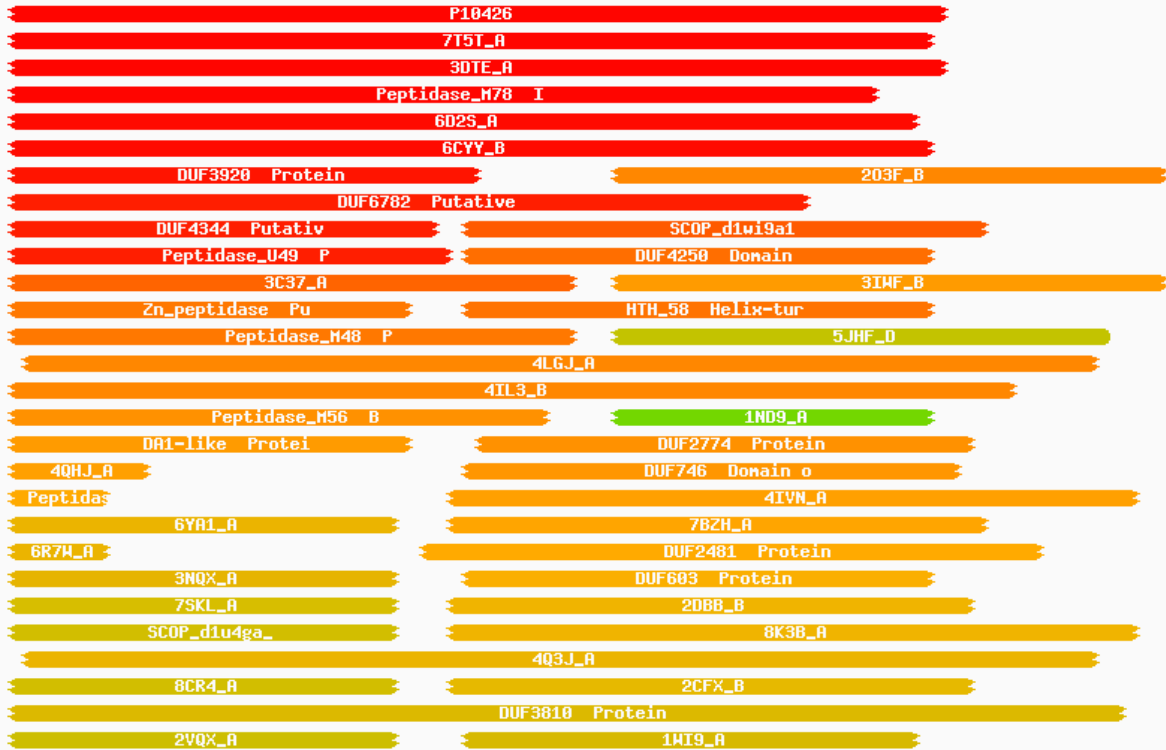


Figure above: Starterator for Gene 52 Sashimi

The short, autoannotated starts (right arrow) and the long starts (left arrow) are conserved across the pham.

HHPred alignments

With the short protein:



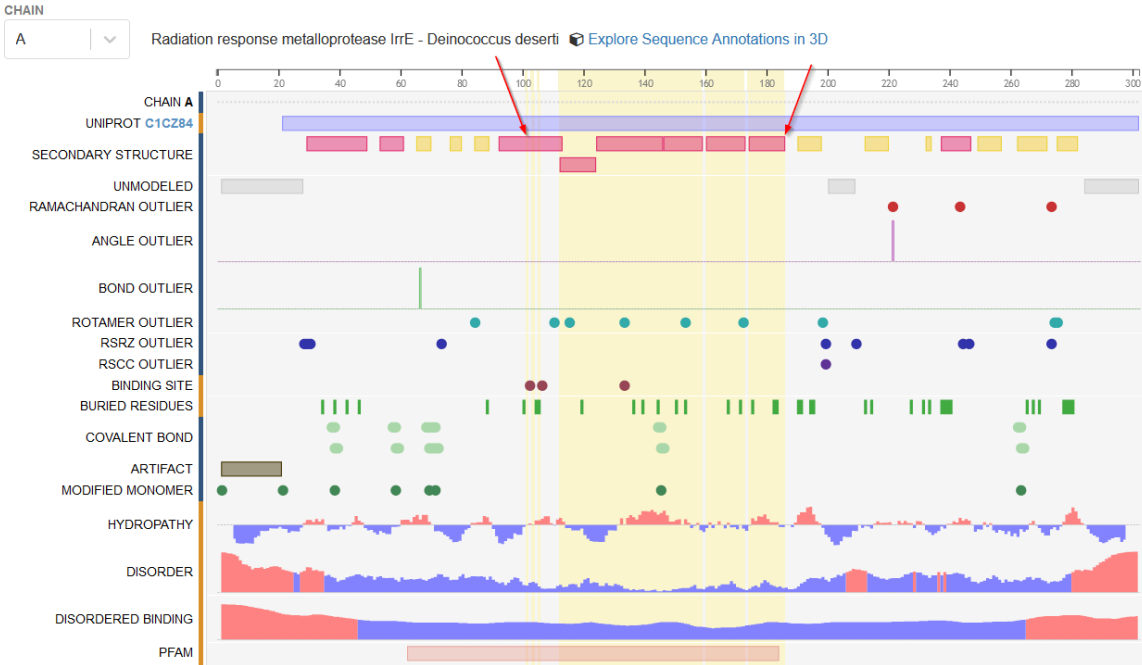
Template alignment | Template 3D Structure | PDBe

3. 3DTE_A IrrE protein; *Deinococcus*, Radiotolerance, Gene regulation, Metallopeptidase, IrrE; HET: MSE; 2.6A (*Deinococcus deserti*)

Probability: 99.22%, E-value: 1.1e-10, Score: 73.95, Aligned cols: 69, Identities: 22%, Similarity: 0.056, Template Neff: 9

```

Q ss_pred          CcchHHHHhCCCC-----CHHHHHHHHHHHhCCCHHHHHHHhC-CC---CHHHHHhCCCCHHHH
Q Raph_50          1  MHELGAHYGHTGV-----TGKQEALNRWAAAYRLVNFDDVLALASV-EQ---SSSGVAAALEVMPDVLET 62 (85)
Q Consensus       1  aHE1~H~~~~~E~a~f~l~i~p~~~~~a~~~~V~~~~~
          |||||++++....          ...|.+.|.+.+.|...+.+. . . .+.++++.|+|...+.
T Consensus      101 aHELGH~11~~~~~E~eAn~FAa~LL~P~~~~~V~~~~~La~~~~vS~~~~ 180 (301)
T 3DTE_A         101 AHEISHALLLGGDDLLSDLHDEYEGDRLEQVIETLCNVGAAALLMPAELIDDLTRFGPTGRALAE LARRADVSA TSALY 180 (301)
T ss_dssp         HHHHHHHHHHCHHHHHHHHHCHHHHHHHHHHHHHHHHHHHHHHHHCHHHHHHHHHHCSHHHHHHHHHHHTCCHHHHH
T ss_pred         HHHHHHHHhCCcchhhhhhhCChHHHHHHHHHHHHHHHHHHhCCHHHHHHHHHHCCCHHHHHHHHHhCCHHHHH
          HHHHHHH
Q ss_pred         HHHHHHH
Q Raph_50         63  YLQLLTR 69 (85)
Q Consensus       63  r~~~~~ 69 (85)
          |+.....
T Consensus      181 R1~~~~~ 187 (301)
T 3DTE_A         181 ALAERTA 187 (301)
T ss_dssp         HHHHTCC
T ss_pred         HHHHCCC
    
```



Aligns with only part of the peptidase domain. The alignment begins with HExxH and ends with the three HTH helices (alignment boundaries shown with arrows). Note the PFAM box at bottom that annotates the metallopeptidase domain.

With the long protein as query:



Template alignment | Template 3D Structure | PDBe

2. **3DTE_A IrrE protein; Deinococcus, Radiotolerance, Gene regulation, Metallopeptidase, IrrE; HET: MSE; 2.6A (Deinococcus deserti)**

Probability: 99.73%, E-value: 1e-15, Score: 103.96, Aligned cols: 107, Identities: 22%, Similarity: 0.153, Template Neff: 9

```

Q ss_pred          CceeEEEECCCCceeEEEECCCCEEEECCCCCHHHHHHHHHHHHHHHHHhC-----ChHHHHHHHHHHH
Q Raph_50long     2  GVRVRTAVTPNGWNGAYDHRRLITLRPGLGPIQLKCTLMHELGHAYGHTGV-----TGKQEALANRWAA 67 (125)
Q Consensus       2  gi~i~~~~~g~~~~~I~i~~~~~l~hE1~H~~~~~-----E~a~~~~~ 67 (125)
                   |||...+++ ..|+...+.|+++...+++|++|+|++++... ..|.+.|||+
T Consensus      63  gi~V~~~~l~~~~g~~~~I~l~~~~r~Ft1aHELGH~l~~~~~-----E~eAn~FAa 141 (301)
T 3DTE_A         63  GITLTFMPMQ-RDGAYDPEHHVILINSQVRPERQRFtLAHEISHALLLGGDDLLSDLHDEYEGDRLEQVIETLCNVGAA 141 (301)
T ss_dssp        63  SCEEEEECTT-CEEEEETTTTEEEEETTSCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
T ss_pred         63  CEEEEeCCC-CeEeCCCCEEEECCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

```



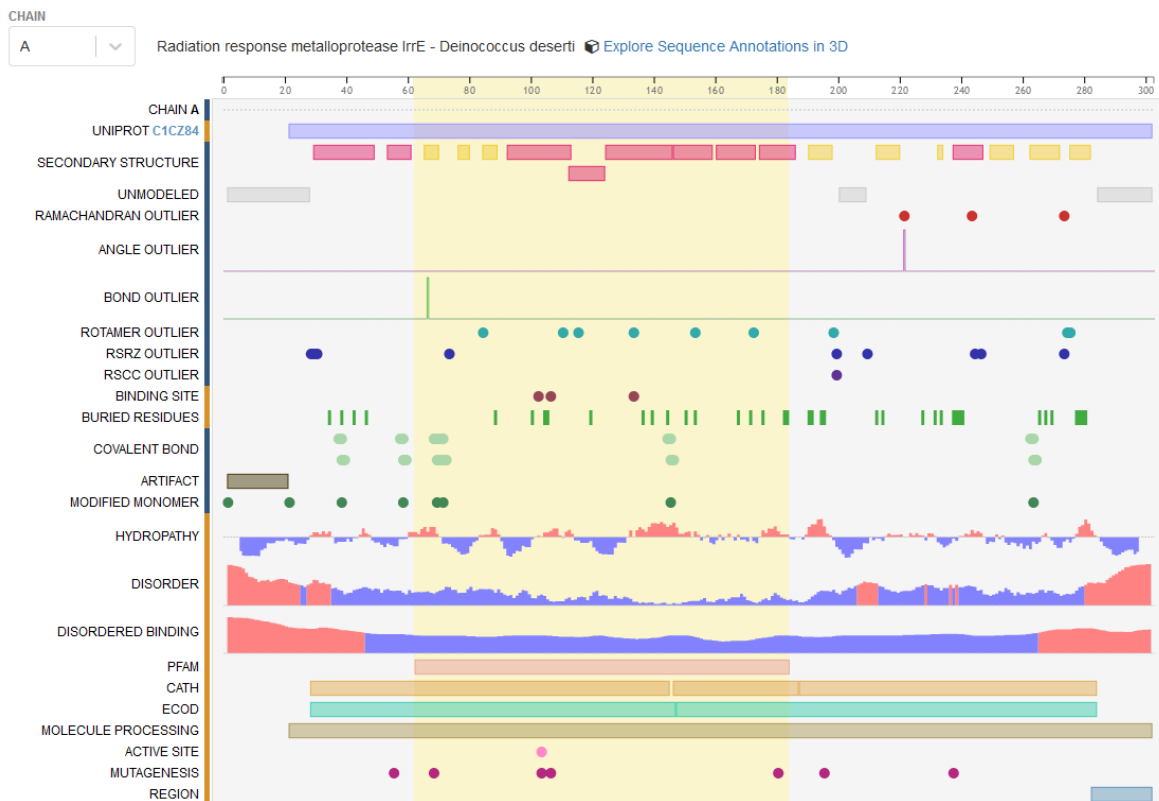
```

Q ss_pred          HHcCHHHHHHHHHhC---CHHHHHhCCHHHHHHHHHHHHH
Q Raph_50long     68  YRLVNFDDVLALASVEQ---SSSGVAAALEVMPDVLETYLQLLTR 109 (125)
Q Consensus       68  ~ll~p~~~~~-----la~~~~V~~~~r~~~~ 109 (125)
                   .+|||...+...+... .+.|+...|+|...+...|+...+
T Consensus      142  ~LL~P~~~~~-----V~L~~~~V~S~~~~R~~~~ 187 (301)
T 3DTE_A         142  ALLMPAELIDDLLTRFGPTGRALAE LARRADV SATSALYALERTA 187 (301)
T ss_dssp        142  HHHSCHHHHHHHHHHHCCSHHHHHHHHHHHCCHHHHHHHHHHTCC
T ss_pred         142  HhhCCHHHHHHHHHHHCCCHHHHHHHHHhCCHHHHHHHHHHHCC

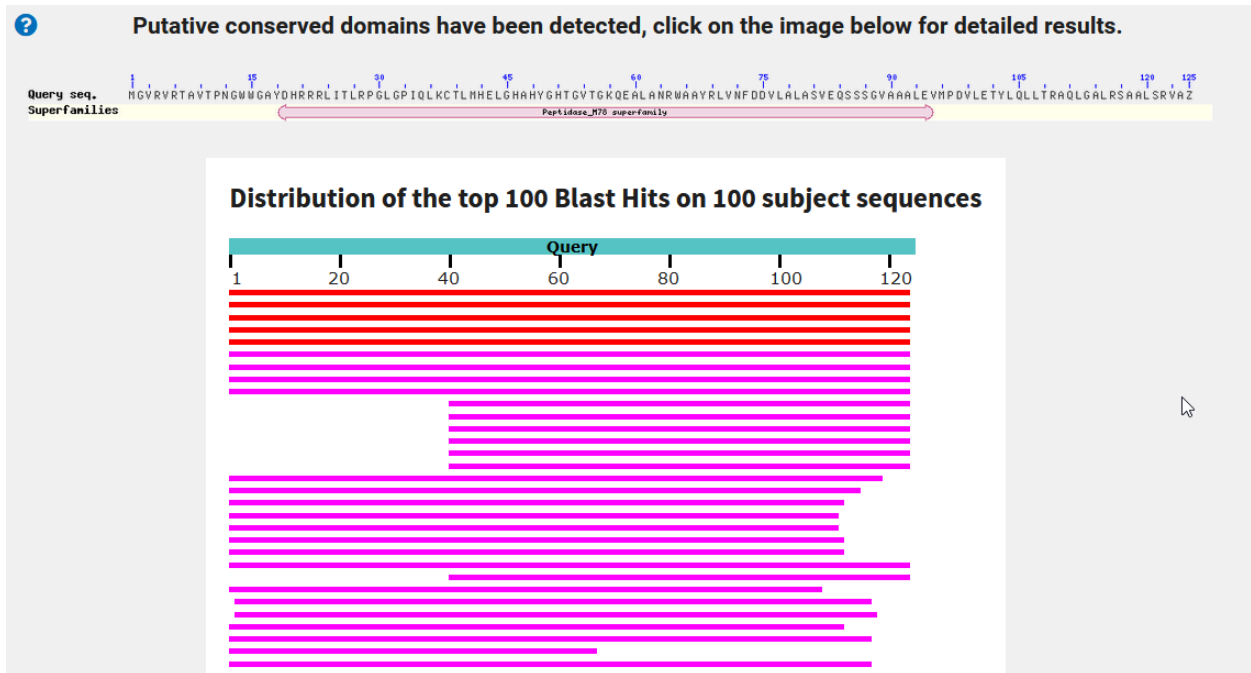
```

Now, the HHPred alignment is across the whole peptidase domain in IrrE

[Help](#)



BLAST and CDD



Aligns across full length of the peptidase M78 domain

Conserved domains on [1cl|Query_1706891] View Standard Results

Local query sequence

Protein Classification

M78 family metallopeptidase domain-containing protein (domain architecture ID 1522)
M78 family metallopeptidase domain-containing protein which includes the catalytic domain of ImmA/IrrE family metallo-endopeptidase

Graphical summary Zoom to residue level show extra options >

Query seq. MGVRVRTAVTPNGMVGAYDHRRLITLRPGLPIQLKCTLMHGLGHAHYGHTGVTGKQALANRWAAAYRLVNFDDVLAASVEQSSSGVAAALEVMPDVLETYLQLLTRAQLGALRSAAALSRVAZ

Specific hits

Non-specific hits

Superfamilies

Peptidase_M78 superfamily

Search for similar domain architectures Refine search

List of domain hits

Name	Accession	Description	Interval	E-value
Peptidase_M78	pfam06114	IrrE N-terminal-like domain; This entry includes the catalytic domain of the protein ImmA, ...	18-95	5.84e-07
ImmA	COG2856	Zn-dependent peptidase ImmA, M78 family [Posttranslational modification, protein turnover, ...	25-68	6.26e-04

Blast search parameters

Data Source: Live blast search RID = 68WF46DF016
User Options: Database: CDSEARCH/cdd Low complexity filter: no Composition Based Adjustment: yes E-value threshold: 0.01 Maximum number of hits: 500

The peptidase domain alignments start at around 20aa or 24aa. The short start proteins start at about 40 aa on this alignment, which cuts out part of the peptidase conserved domain.

See Official Function List for naming. These examples hit IrrE-like metallopeptidases, which include a HTH domain.

DNA alignment shows sequence and reading frame conservation in the gap with no coding potential

ClustalW close-up view, like Starterator

DNA sequences: Relatives of MidnightRain_55, plus 500 bp downstream of the stop codon, which includes the full length of the downstream gene, MidnightRain_54 and its relatives.

CLUSTAL 2.1 multiple sequence alignment

```
AbbyDaisy_AY_gp51      GCTACCGCAGAACCCCTAGAGTTCTTGCCCAACAGATGGGCGTGAGGGTT
Auxilium_AY_gp48      GCTACCGCAGAACCCCTAGAGTTCTTGCCCAACAGATGGGCGTGAGGGTT
Seahorse_AY_gp55      GCTTCCACACGAACCCTAGAGTTCTCGCTCAGCAGATGGGCGTGAGGGTT
Isolde_AY_gp51       GCTTCCACACGAACCCTAGAGTTCTCGCTCAGCAGATGGGCGTGAGGGTT
Raphaella_AY_gp51    GCTTCCACACGAACCCTAGAGTTCTCGCTCAGCAGATGGGCGTGAGGGTT
YoungHarleezy_Draft_AY_gp52 GCTTCCACACGAACCCTAGAGTTCTCGCTCAGCAGATGGGCGTGAGGGTT
Aikyam_Draft_AY_gp50 GCTTCCACACGAACCCTAGAGTTCTCGCTCAGCAGATGGGCGTGAGGGTT
CookieBear_Draft_AY_gp52 GCTTCCACACGAACCCTAGAGTTCTCGCTCAGCAGATGGGCGTGAGGGTT
Gorpy_AY_gp52        GCTTCCACACGAACCCTAGAGTTCTCGCTCAGCAGATGGGCGTGAGGGTT
Sakai_AY_gp51       GCTTCCACACGAACCCTAGAGTTCTCGCTCAGCAGATGGGCGTGAGGGTT
Richie_AY_gp54      GCTTCCACACGAACCCTAGAGTTCTCGCTCAGCAGATGGGCGTGAGGGTT
Tiff81_Draft_AY_gp52 GCTTCCACACGAACCCTAGAGTTCTCGCTCAGCAGATGGGCGTGAGGGTT
Sashimi_Draft_AY_gp53 GCTACCGCAGAACCCCTAGAGTTCTTGCCCAACAGATGGGCGTGAGGGTT
Globfish_Draft_AY_gp52 GCTTCCACACGAACCCTAGGGTTCTCGCTCAGCAGATGGGCGTGAGGGTT
BenchScraper_Draft_AY_gp50 GCTACCGCAGAACCCCTAGAGTTCTTGCCCAACAGATGGGCGTGAGGGTT
MidnightRain_AY_gp55 GCTACCGCAGAACCCCTAGAGTTCTTGCCCAACAGATGGGCGTGAGGGTT
BillyTP_Draft_AY_gp53 GCTACCGCAGAACCCCTAGAGTTCTTGCCCAACAGATGGGCGTGAGGGTT
Raqqa_Draft_AY_gp43  GCTTCCACACGAACCCTAGAGTTCTCGCTCAGCAGATGGGCGTGAGGGTT
Hillester_Draft_AY_gp55 GCTTCCACACGAACCCTAGAGTTCTCGCTCAGCAGATGGGCGTGAGGGTT
RadFad_AY_gp55      GCTTCCACACGAACCCTAGAGTTCTCGCTCAGCAGATGGGCGTGAGGGTT
Faja_AY_gp52       GCTACCGCAGAACCCCTAGAGTTCTTGCCCAACAGATGGGCGTGAGGGTT
Phrank15_Draft_AY_gp53 GCTACCGCAGAACCCCTAGAGTTCTTGCCCAACAGATGGGCGTGAGGGTT
Anekin_Draft_AY_gp49 GCTACCGCAGAACCCCTAGAGTTCTTGCCCAACAGATGGGCGTGAGGGTT
Persistence_AY_gp47 GCTACCGCAGAACCCCTAGAGTTCTTGCCCAACAGATGGGCGTGAGGGTT
*** ** ***** ** * * ***** ** * * ***** ** * *
```

Stop, upstream gene "long start"

```
AbbyDaisy_AY_gp51      CGGACGGCTGTACACCTAACCGGATGGTGGGGCGCGTACGATCACAGGCG
Auxilium_AY_gp48      CGGACGGCTGTACACCTAACCGGATGGTGGGGCGCGTACGATCACAGGCG
Seahorse_AY_gp55      AGAACAGCTGTACACCGCATGGATGGTGGGGCGCATACGATCATCGGCG
Isolde_AY_gp51       CGAACAGCCGTGACCCCTAATGGATGGTGGGGCGCTTACGATCACAGGCG
Raphaella_AY_gp51    CGAACAGCCGTGACCCCTAATGGATGGTGGGGCGCATACGATCACAGGCG
YoungHarleezy_Draft_AY_gp52 CGAACAGCCGTGACCCCTAACCGGATGGTGGGGCGCATACGATCACAGGCG
Aikyam_Draft_AY_gp50 CGAACAGCTGTACACCGAATGGTGGTGGGGCGCGTACGATCACAGGCG
CookieBear_Draft_AY_gp52 CGAACAGCCGTGACCCCTAACCGGATGGTGGGGTGCATAACGATCACAGGCG
Gorpy_AY_gp52        CGAACAGCCGTGACCCCTAATGGATGGTGGGGCGCATACGATCACAGGCG
Sakai_AY_gp51       CGAACAGCCGTGACCCCTAATGGATGGTGGGGCGCATACGATCACAGGCG
Richie_AY_gp54      CGAACAGCCGTGACCCCTAACCGGATGGTGGGGTGCATAACGATCACAGGCG
Tiff81_Draft_AY_gp52 CGAACAGCTGTACACCGCATGGATGGTGGGGCGCGTACGATCACAGGCG
Sashimi_Draft_AY_gp53 CGGACAGCTGTGACCGCGCATGGTGGTGGGGCGCGTACGATCACAGGCG
Globfish_Draft_AY_gp52 CGGACGGCTGTACACCTAACCGGATGGTGGGGCGCGTACGATCATCGGCG
BenchScraper_Draft_AY_gp50 CGGACGGCTGTACACCTAACCGGATGGTGGGGCGCGTACGATCACAGGCG
MidnightRain_AY_gp55 CGGACGGCTGTTACACCTAACCGGATGGTGGGGCGCGTACGATCACAGGCG
BillyTP_Draft_AY_gp53 CGAACAGCTGTACACCTAACCGGATGGTGGGGCGCGTACGATCACAGGCG
Raqqa_Draft_AY_gp43  CGAACAGCCGTGACCCCTAATGGATGGTGGGGCGCTTACGATCACAGGCG
Hillester_Draft_AY_gp55 CGAACAGCTGTACACCGAATGGTGGTGGGGCGCGTACGATCACAGGCG
RadFad_AY_gp55      CGAACAGCTGTACACCGAATGGTGGTGGGGCGCGTACGATCACAGGCG
Faja_AY_gp52       CGAACAGCTGTACACCTAACCGGATGGTGGGGCGCGTACGATCACAGGCG
Phrank15_Draft_AY_gp53 CGAACAGCTGTACACCTAACCGGATGGTGGGGCGCGTACGATCACAGGCG
Anekin_Draft_AY_gp49 CGGACGGCTGTACACCTAACCGGATGGTGGGGCGCGTACGATCACAGGCG
Persistence_AY_gp47 CGGACGGCTGTACACCTAACCGGATGGTGGGGCGCATACGATCACAGGCG
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
```

