Genome Evolution:
Duplication (Paralogs) & Degradation (Pseudogenes)
Duplication and Degradation Module

- Module Instructions

**Paralog**

go to the IMG Gene Details page for the proposed gene

Gene product name

- name

Percent identity

- identity

Alignment length

- length

E-value

- e-value

Pairwise alignment

- alignment

**Pseudogene**

- [http://pfam.sanger.ac.uk](http://pfam.sanger.ac.uk)

Is this a pseudogene?

- enter in lab report

- Genes related by duplication within a genome
- May evolve new function
- Non-functional
- Share homology with functional gene
Sequence Conservation Reflects Evolutionary Relationships (Ancestry)

• Homologs
  – Orthologs
    • Genes duplicated via appearance of new species
      – Identical function in different organisms
  – Paralogs
    • Genes duplicated within a species
      – Perform slightly different tasks in cell
        » Can develop new capabilities
        » Can become pseudogene if functionality lost but sequence similarity retained
Let's hunt for paralogs first. . .
Go to Gene Detail page for your gene

### Gene Detail

#### Gene Information

<table>
<thead>
<tr>
<th>Gene Information</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene Object ID</td>
<td>2500607069</td>
</tr>
<tr>
<td>Gene Symbol</td>
<td></td>
</tr>
<tr>
<td>Locus Tag</td>
<td>PlimDRAFT_19450</td>
</tr>
<tr>
<td>Product Name</td>
<td>Uncharacterized anaerobic dehydrogenase, COG3383</td>
</tr>
<tr>
<td>IMG Product Source</td>
<td>COG3383</td>
</tr>
<tr>
<td>Genome</td>
<td>Planctomyces limnophilus DSM 3776</td>
</tr>
<tr>
<td>DNA Coordinates</td>
<td>2111568..2113793 (-)(2226bp)</td>
</tr>
<tr>
<td>Scaffold Source</td>
<td>Planctomyces limnophilus DSM 3776 : PlimDRAFT 4083246 C168 (5423025bp)</td>
</tr>
<tr>
<td>IMG ORF Type</td>
<td></td>
</tr>
<tr>
<td>GC Content</td>
<td>0.58</td>
</tr>
<tr>
<td>External Links</td>
<td></td>
</tr>
<tr>
<td>Fused Gene</td>
<td>No</td>
</tr>
<tr>
<td>Fusion Component</td>
<td>No</td>
</tr>
</tbody>
</table>

OID 2500607069
Homolog Display

Customized Homolog Display

Select Paralogs/Orthologs from drop-down menu

Version 2.6, August 2008
Questions/Comments
©2008 The Regents of
Disclaimer
Tables displaying orthologs and paralogs

Gene Homolog

Homolog Display

Customized Homolog Display

- Homolog Selection: Paralogs / Orthologs

- Include current query gene 2500607069 in this page for adding to gene cart.

Paralogs

Paralogy are reciprocal hits within the same genome.

<table>
<thead>
<tr>
<th>Select</th>
<th>Paralog</th>
<th>Product Name</th>
<th>Percent Identity</th>
<th>Alignment On Query Gene</th>
<th>Alignment On Subject Gene</th>
<th>Length</th>
<th>E-value</th>
<th>Bit Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2500608990</td>
<td>oxidoreductase alpha (molybdopterin) subunit</td>
<td>22.85</td>
<td></td>
<td></td>
<td>726aa</td>
<td>2.0e-23</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td>2500606317</td>
<td>NADH dehydrogenase subunit G (EC 1.6.5.3) (IMQtern)</td>
<td>30.69</td>
<td></td>
<td></td>
<td>569aa</td>
<td>9.0e-03</td>
<td>45</td>
</tr>
</tbody>
</table>

Add Selections To Gene Cart  Select All  Clear All  Phylogenetic Distribution

Orthologs

(Orthologs are bidirectional best hits from BLASTP of each genome against each other genome.)

Domains(D): B=Bacteria, A=Archaea, F=Fukarya, P=Plasmids, V=Viruses.

Genome Completion(C): F=Finished, D=Draft.

Click on column name to sort.

<table>
<thead>
<tr>
<th>Select</th>
<th>Ortholog</th>
<th>Product Name</th>
<th>Percent Identity</th>
<th>Alignment On Query Gene</th>
<th>Alignment On Subject Gene</th>
<th>Length</th>
<th>E-value</th>
<th>Bit Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>641109856</td>
<td>nitrate reductase catalytic subunit</td>
<td>67.86</td>
<td></td>
<td></td>
<td>732aa</td>
<td>0.0e+00</td>
<td>1052</td>
</tr>
<tr>
<td></td>
<td>64132919</td>
<td>Nitrate reductase</td>
<td>65.53</td>
<td></td>
<td></td>
<td>757aa</td>
<td>0.0e+00</td>
<td>991</td>
</tr>
<tr>
<td></td>
<td>642471882</td>
<td>molybdopterin oxidoreductase</td>
<td>62.26</td>
<td></td>
<td></td>
<td>744aa</td>
<td>0.0e+00</td>
<td>965</td>
</tr>
<tr>
<td></td>
<td>638980649</td>
<td>nitrate reductase catalytic subunit</td>
<td>63.93</td>
<td></td>
<td></td>
<td>730aa</td>
<td>0.0e+00</td>
<td>949</td>
</tr>
</tbody>
</table>

We will revisit the ortholog table to complete last of the modules next week.

For those cases in which no paralogs are displayed, enter “No paralogs found” in your notebook.

For this module, focus on paralog table.
For each paralog, perform a reciprocal BLAST search with the amino acid sequence for the paralog as your query against the *P. limnophilus* genome. Then inspect the alignment of this paralog with our assigned gene → notebook.

<table>
<thead>
<tr>
<th>Select</th>
<th>Paralog</th>
<th>Product Name</th>
<th>Percent Identity</th>
<th>Alignment On Query Gene</th>
<th>Alignment On Subject Gene</th>
<th>Length</th>
<th>E-value</th>
<th>Bit Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐</td>
<td>2500608990</td>
<td>oxidoreductase alpha (molybdopterin) subunit</td>
<td>22.85</td>
<td></td>
<td></td>
<td>726aa</td>
<td>2.0e-23</td>
<td>114</td>
</tr>
<tr>
<td>☐</td>
<td>2500606137</td>
<td>NADH dehydrogenase subunit G (EC 1.6.5.3) (IMGterm)</td>
<td>30.69</td>
<td></td>
<td></td>
<td>569aa</td>
<td>9.0e-03</td>
<td>45</td>
</tr>
</tbody>
</table>

These are the statistics you record in your notebook.
Right-click on the paralog OID; open Gene Detail page in new tab

<table>
<thead>
<tr>
<th>Select</th>
<th>Paralog</th>
<th>Product Name</th>
<th>Percent Identity</th>
<th>Alignment On Query Gene</th>
<th>Alignment On Subject Gene</th>
<th>Length</th>
<th>E-value</th>
<th>Bit Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2500608990</td>
<td>oxidoreductase alpha (molybdopterin) subunit</td>
<td>22.85</td>
<td></td>
<td></td>
<td>726aa</td>
<td>2.0e-23</td>
<td>114</td>
</tr>
</tbody>
</table>

Gene Information

<table>
<thead>
<tr>
<th>Gene Information</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene Object ID</td>
<td>2500608990</td>
</tr>
<tr>
<td>Gene Symbol</td>
<td>PlimDRAFT_38660</td>
</tr>
<tr>
<td>Locus Tag</td>
<td>PlimDRAFT_38660</td>
</tr>
<tr>
<td>Product Name</td>
<td>oxidoreductase alpha (molybdopterin) subunit</td>
</tr>
<tr>
<td>IMG Product Source</td>
<td>TIGR01701</td>
</tr>
<tr>
<td>Genome</td>
<td>Planctomycetes limnophilus DSM 3776</td>
</tr>
<tr>
<td>DNA Coordinates</td>
<td>4322085..4324265 (+)(2181bp)</td>
</tr>
<tr>
<td>Scaffold Source</td>
<td>Planctomycetes limnophilus DSM 3776 : PlimDRAFT 4083246_C168 (5423025bp)</td>
</tr>
<tr>
<td>IMG ORF Type</td>
<td></td>
</tr>
<tr>
<td>GC Content</td>
<td>0.55</td>
</tr>
<tr>
<td>External Links</td>
<td></td>
</tr>
<tr>
<td>Fused Gene</td>
<td>No</td>
</tr>
<tr>
<td>Fusion Component</td>
<td>No</td>
</tr>
</tbody>
</table>

Protein Information

<table>
<thead>
<tr>
<th>Amino Acid Sequence Length</th>
<th>726aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>COG</td>
<td>COG0243 - Anaerobic dehydrogenases, typically selenocysteine-containing</td>
</tr>
</tbody>
</table>
Use the paralog sequence for reciprocal BLAST search

COPY the entire protein sequence in FASTA format then go to “Find Genes”
Select BLAST from menu

Paste Protein or DNA sequence here:

```
>2500638990 oxidoreductase alpha (mooybdopterin) subunit [Planctomyces limnophilus DSM 3776 : PlimDRAFT_4083246_C166]
MRKITTGGGP4VLYTLRKNEMGGFKSWMQAMRSKNACK7CALGMGGQK
GGVNEAGSFPEVCKSLQAMASDLQAGIRSEFWQKTSIGQKLKSPLREL
EYCGRLVEPVHEAGK5HYRPK5MEAFARIAKL5VQGDTEFY5GR
SSNEAGFLQL PARA1YGTNNINMCPSFQCHQS35VGLTSVLGTGTATLTL
DVEHACCFILGFLGAFSNHPRLMSTLKHRRRGGEVIVINVPETGLVNF
SVFDSPWSSFLGKTIALSTYVQPIGGDLALLGIAKRIELGQHDPALFL
TACGWSWEGKHSLESTHUSEIECEKSCGосEINAIARREYAESKHNTVFQAUT
MGITHHAGVENVEAIANLIMRGMVGRPHAGLMPIRHERSN1IQGMGTGVV
TPKLDUWFERLQSTQISLPQTPGRDTMACMDGSLNRELEFGFCCLGGLNL
```

Program: blastp (Protein vs. Protein)
E-value: 1e-2

Databases:
- Planctomyces limnophilus DSM 3776 (B)[D]
- Planctomyces maris DSM 8797 (B)[D]
- Plesiocystis pacifica SIR-1 (B)[D]
- Polaribacter igensii 23-P (B)[D]
- Polaromonas naphthalenivorans CJ2 (B)[F]
- Polaromonas sp. JS866 (B)[F]
- Polynucleobacter necessarius STIR1 (B)[F]
- Polynucleobacter sp. QLW-P1DMWA-1 (B)[F]
- Porphyromonas gingivalis W83 (B)[F]
- Prochlorococcus marinus A89601 (B)[F]
- Prochlorococcus marinus MIT9311 (B)[F]

Notes on BLAST database genomes:
- Domains: (B)acteria, (A)rchaea, (E)ukarya, (P)lasmids, (V)iruses.
- Genome Completion: [F]inished, [D]raft.

Select P. limnophilus as database
Press “Run BLAST”
BLASTP 2.2.15 [Oct-15-2006]


Query= 2500606990 oxidoreductase alpha (molybdopterin) subunit [Planctomyces limnophilus DSM 3776 : PlimDRAFT_4083246_C168] (726 letters)

Database: /home/img/ezeto/img_dev/v2/dataLoad/web-data/taxon.faa/2500575009.faa.Elastdb/2500575009 4891 sequences; 1,566,052 total letters

Searching.........................................................done

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>Score</th>
<th>E</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2500606990</td>
<td>PlimDRAFT_36660 oxidoreductase alpha (molybdopterin) ...</td>
<td>1432</td>
<td>Oe+00</td>
<td></td>
</tr>
<tr>
<td>2500607069</td>
<td>PlimDRAFT_19450 Uncharacterized anaerobic dehydrogena...</td>
<td>105</td>
<td>9e-24</td>
<td></td>
</tr>
</tbody>
</table>

Find the Gene OID in hit list that corresponds to your assigned gene

Inspect the pair-wise alignment

Scroll down
Reciprocal BLAST search results

>2500607069 P1imDRAFT_19450 Uncharacterized anaerobic dehydrogenase
[Planctomyces limnophilus DSM 3776]
Length = 741

Score = 105 bits (263), Expect = 9e-24
Identities = 151/674 (22%), Positives = 250/674 (37%), Gaps = 80/674 (11%)

Query: 98 FELEYCGRLVEPVIHEAGKSHYRPIISWEAFARIARLKSVO----PDETWFYFSGRSSN 153
R L+ R P+ P+ + A + R + + Q P + + + +
Sbjct: 97 FVLADAPDRATTPRRAFDGRLEPVVDAAALFLGRTGRFQIQEQHGPFSVAFISTQGMP 156

Query: 154 EAGFLQLFLARI-YGTNNINNC5YCHQAXXAXXXXXXXXEDVEHADCFGL 212
E L L A+ + C + + + D E + D + + G
Sbjct: 157 EEMALLGAVAKGFMRLHGODGNTQCHNATAVAYKEAFGFDAPPPITYADFEESDAMVFVG 216

Query: 213 GNPASNHPLMLSTLKHIERRGREGGEVIVINPVETGLVNFSVPSDWSLLFGETKIASLYVQP 272
NP H P + + RR E+IV++P + + T+ L +QP
Sbjct: 217 SNPCIAHPIWVMR-NRSSPEIIIVDP-----------RGWETTVYATQ--HLAQIP 261

Query: 273 QIGGDIALLLTGIARLELGQHDPALFTTACDGUWQHLESTHSEICEKRSVGGIDEI 332
+ D L G+A+ ++E G+ D P+ + G E+ + + + ++G+ ++I
Sbjct: 262 KT--DQTLFYQVARLLLIEQGRIDEKFVLESTTGEEFARFVNVDSLARTASETGLEPNQ 319

Query: 333 NAIARRYASKNTFXWMTGMITHAHHVENVEAIANLMLRMHGGVPIAHLPRIHGSHNI 392
A + ++ F WMG+ GV ++I NLAM G +G+P G I G N
Sbjct: 320 EFACHTXKEXRVSFWWNTGEXQHQPVRATQSIINLALMNTGNIKGPGCOAStITGQCN 379

Query: 393 QGMGTYGVTXKLI---------KDIVFELLQSTFQISLPQTPGRDTMDAGCDGSNLRELKF 442
G T L + S + +P M+G L +K
Sbjct: 380 KGSRLELSNNTNLLQGHDKNADHRKIAGLISMDGDSVITPENSWSYHEIEMICILKEIKG 439

Query: 443 GFCGLGNNLFGSNPDAAAYAALSKLDQIVYLYSTTINTGHGALQETILLPULARDEPE 502
+ + N S + A +LS+LD +V NT A + +LP E E
Sbjct: 440 LWIICNTNTAHSWIONQLEMLLSDLDLFLVQYDMYNTETAQ--MADLVLPAAGWEGK-E 495

Query: 503 TTTQESBMNYMRBDGGPDRTG-PLSEIHVIAEIQ------------------------RIV 540
T S R+ + R G L++ H+ +A+
Sbjct: 496 GTFINSERRIGRIK--VRRAPGKALADPHIFQQLVAEYWGCGEQFRUSPEAVFEVLKE 553

Copy/paste the alignment into your notebook
Recording results in your notebook

Paralog

go to the IMG Gene Details page for the proposed gene

Gene OID

2501576450

Gene product name

oxidoreductase alpha (molybdopterin) subunit

Percent identity

22.85

Alignment length

726aa

E-value

2.0e-23

Pairwise alignment

Query: 98 PELEYCGRVVEEVHEAGKSHYRPISWEDAFARIAERLKSVO----EDETFWYFSGRSSN 153
       R I+ R P+ P+ W+ A R + TQ F + +G+

Sbjct: 97 RVLDAAPDRATTELRYADKNRLEPVDWDAALRTFTGRFRQQI5KHGEH3VAFISTGQMEPT 156

Enter the gene OID, gene name, and alignment statistics from Gene Detail page for paralog

(Will need to add heading/box for OID)
Recording results in your notebook

Remember, the alignment generated by reciprocal BLAST search is between the paralog (query) and your assigned gene (subject).

Repeat process for all paralogs with significant E-value.
PSEUDOGENES

What are they?

- Genes that are nonfunctional.
- Will align well to known protein sequence on BLAST, CDD, and/or Pfam and may appear to encode a legitimate coding sequence (start/stop codons, Shine-Dalgarno sequence within proper distance, etc.)

- Formed by one of two mechanisms:
  - By duplication of functional gene followed by mutagenesis that removes functionality
  - Degradation of a functional gene no longer required by organism
If an ORF meets one of the following criteria, then it should be annotated as a possible pseudogene.

- Sequence is interrupted by more than one stop codon or frameshift, corresponds to a truncated Pfam less than 30% of predicted profile.

- Sequence separated by another ORF

- Missing key residues known to be required for functionality.

New resource: ScanProsite
For this class, we will investigate possible pseudogenes using only criterion 1 and criterion 3. We do not have the resources to obtain sequence information needed for criterion 2; however, a brief explanation will be provided.

Some things to keep in mind:

✓ pseudogenes are very RARE

✓ the criteria used to characterize pseudogenes is based on computational methods → experimental confirmation is required before a gene can be ruled nonfunctional
CRITERION #1

- Sequence is interrupted by more than one stop codon or frameshift, corresponds to a truncated Pfam less than 30% of predicted profile

Navigate to Pfam database

Pseudogene

http://expasy.org/tools/scanprosite
http://pfam.sanger.ac.uk
Is this a pseudogene?

Enter in Lab Report.

Quick Links
- Sequence Search: Analyze your protein sequence for Pfam matches
- View a Pfam Family: View Pfam family annotation and alignments
- View a Clan: See groups of related families
- View a Sequence: Look at the domain organisation of a protein sequence
- View a Structure: Find the domains on a PDB structure
- Keyword Search: Query Pfam by keywords
- Jump To: Enter any accession or ID to jump to the page for a Pfam family or clan, UniProt sequence, PDB structure, etc.
CRITERION #1

- Copy/paste the amino acid sequence in FASTA format for your assigned gene into the query box
- Click “Go”
CRITERION #1

- On results page, note the domain graphic

Not a pseudogene:

- If the last domain is truncated and running to the end of the sequence, then one should investigate the possibility that this ORF is a pseudogene – HOW?
CRITERION #1

Possible pseudogene:

- Determine whether this domain is required for functionality of the protein
  1. Calculate length of sequence aligned to HMM
     EX: \((113 - 1) + 1 = 113\) residues

Significant Pfam-A Matches
Show or hide all alignments.

<table>
<thead>
<tr>
<th>Pfam-A</th>
<th>Description</th>
<th>Entry type</th>
<th>Sequence</th>
<th>HMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molybdcop_Fe4S4</td>
<td>Molybdopterin oxidoreductase Fe4S4 domain</td>
<td>Domain</td>
<td>47 101</td>
<td>1 63</td>
</tr>
<tr>
<td>Molybdopterin</td>
<td>Molybdopterin oxidoreductase</td>
<td>Family</td>
<td>104 531</td>
<td>1 499</td>
</tr>
<tr>
<td>Molybd binding</td>
<td>Molybdopterin dinucleotide binding domain</td>
<td>Domain</td>
<td>627 733</td>
<td>1 113</td>
</tr>
</tbody>
</table>
2. Determine the length of the HMM model
   EX: 113

2. Calculate % coverage:
   Divide the value from step 1 by the value from step 2 and multiply by 100
   EX: 113 / 113 x 100 = 100%

If this value is < 30%, then this may be a pseudogene. Research (PubMed) must indicate that the domain is required for functionality before one can conclude it is a pseudogene – look it up!
CRITERION #2

- Sequence separated by another ORF

Possible pseudogene:

- No tools available to easily do this on img/edu, but this is how it is done in theory

1. Obtain genomic DNA sequence that is flanking your ORF (1000s of kilobases on one side of your gene or the other)
2. Perform Pfam search
3. Note the domain graphic

- If the second half of the fragmented domain is present in the flanking DNA, and the domain is required for functionality (consult PubMed!), then this may be a pseudogene.
CRITERION #3

- Missing key residues known to be required for functionality.

Possible pseudogene:

- Navigate to ScanProsite tool on Prosite database

Pseudogene

http://expasy.org/tools/scanprosite
http://pfam.sanger.ac.uk
Is this a pseudogene?

Enter in Lab Report.
New online resource!

http://expasy.org/tools/scanprosite

ScanProsite

What is it?

• curated database of multiple sequence alignments of motifs used for the purpose of identifying domains and families of protein sequences
  ✓ alignments are known as profiles
  ✓ similar to Pfams or COGs

What does it do?

• generally used to identify an open reading frame as a pseudogene by verifying the absence of catalytic residues
• also can be used to identify proteins present in distantly related microbes
Copy/paste the amino acid sequence in FASTA format for your assigned gene into the query box.

- Deselect “Exclude motifs with high probability of occurrence”
- Check the box “Show low level score”
- Press “START THE SCAN”
ScanProsite Results

found: 32 hits in 1 sequence

2500607069 (741 aa)

MSWYQDNPIMQOPQTLLHOREGLTTRALQQPAGFGLQLPQSLPTUIATTDNYCGFCSTGCGLKVF
LKEGERGGLSFSTLSYFVNLGNACPKGHMCALRYVDAFDRATTFLYRAEKNRLPEVVDVDAALRLUFTG
RFQIQEKHGPHSVAPFISTQGMTEEMALLGAVAKFCMGRLHGDGNTROCMATAVVAYKEAAGFDA
PPYYTADFEESDAMVFGSNCPLAIHHIMDERVHMDNSPEIIVVDPEGNETMYAT0HLAGP4KTD
QTLFYGVARLLIIQGRIDEKFVLSTSTTFEEFAFVFDVLYSLARTASETLEGPMQIERFACTHEKE
RVSFLWVTGVNQSHOGVTRAO8ISINLALMTGNIKPGTANSTGQCNAMGSRMLFSETTNLLGHD
FKWADHRKAIACILSMDESVIPTENSW3YHEIMECGILKEIKGLWI1CTW1H3SWIQNLAEMLS
RLDFLYVQDMHYITETAAQMAZLVLPAAGWGEKECTFINERRGIRIHKVRRAPGKALADPHIFQLV
AEGYSECCEQFGRSSPEAVFVEVKECSRFCDPTGQVNYRQIEEQQG6VQWCPEGHTAVQVENRRL
FEDFCYFHYADGMRAKFLFA5LSRPITLPEAYFPILLTGRGSA5QWHQ7TRTAKS5PF1KRLY?ERPF
IXEPADARALKIAHNAWYVE3KRGPSRAMALTVPFAVRQGYF1AMYEG7NQLTDAYF?YSKQP
PSTKACAVNVK6AE

Query sequence used for search

ruler:

| 1 | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 |

hits by patterns with a high probability of occurrence or by user-defined patterns: [32 hits (by 6 distinct patterns) on 1 sequence]

Visual representation of profiles and patterns

Portion of sequence aligning to profile
(If scroll over, will highlight corresponding sequence in query above)
Analysis of ScanProsite Results – Do I have a pseudogene by Criterion 3?

- Inspect the results
  - Underlined series of residues, which correspond to required features in a functional protein domain

**PS5022  EF_HAND_2  EF-hand calcium-binding domain profile:**

152 - 187:  score = 8.293

QIEFWLRKQFYSVDNHEDRISAKDLOKMLSGqNYR

Predicted feature:

<table>
<thead>
<tr>
<th>CA_BIND</th>
<th>165</th>
<th>176</th>
<th>(Potential)</th>
</tr>
</thead>
</table>

[condition: D-x-[DNS]-[LVFWY]-[DENSTG]-[DNQGKRX]-[GR]-[ILVMC]-[DENQSTAGC]-x(?)-[DE]]

- Red colored residues, which correspond to necessary components of an active site

**PS50007  PIPLC_X_DOMAIN  Phosphatidylinositol-specific phospholipase X-box domain profile:**

320 - 464:  score = 60.694

DTMNNPLSHYWISSIHHTYTLGDQFSSESESSLAYARCLRMGCRCIELDCWDSPDGMFYIV

<table>
<thead>
<tr>
<th>ACT_SITE</th>
<th>335</th>
<th>By similarity</th>
<th>[condition: H]</th>
<th>[group: 1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT_SITE</td>
<td>380</td>
<td>By similarity</td>
<td>[condition: H]</td>
<td>[group: 1]</td>
</tr>
</tbody>
</table>
Analysis of ScanProsite Results –
Do I have a pseudogene by Criterion 3?

- If there are **no** underlined sequences or red residues, then the protein is missing the predicted features or active site components required for functionality.

Although this is strong evidence that your gene is a pseudogene, you must confirm that there are no exceptions to a condition for functionality – HOW?
Click on **PS***** identification number in the title line for the profile in which the condition for functionality has not been met**

<table>
<thead>
<tr>
<th>Identification</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS00008</td>
<td>MYRISTYL N-myristoylation site</td>
</tr>
<tr>
<td>74 - 79</td>
<td>GLspTT</td>
</tr>
<tr>
<td>152 - 157</td>
<td>GQmpTE</td>
</tr>
<tr>
<td>171 - 176</td>
<td>GLrhGD</td>
</tr>
<tr>
<td>216 - 221</td>
<td>G3npCI</td>
</tr>
</tbody>
</table>

On profile page, look for blue box in “Technical section”

- If multiple sequences are detected in Swiss-Prot that match the profile, and the notes indicate there are exceptions to the conditions, then you cannot conclude you have a pseudogene

---

**Technical section:**

PROSITE method (with tools and information) covered by this documentation:

- **MYRISTYL, PS00008; N-myristoylation site** (PATTERN with a high probability of occurrence)

  **Consensus pattern:**  
  G - {EDRKHPFYW} - x(2) - [STAGCN] - {P} [G is the N - myristoylation site]

- Scan Swiss-Prot/TrEMBL entries against PS00008
- view ligand binding statistics

---

Exceptions
If multiple sequences are detected in Swiss-Prot that match the profile, and there are NO exceptions to the conditions, then you hypothesize that you have a pseudogene.
Recording results in your notebook

If you cannot conclude you have a pseudogene, enter “NO” in the box.

If you can conclude you may have a pseudogene, enter “YES” in the box with a brief note as to which criterion was used to come to this conclusion.