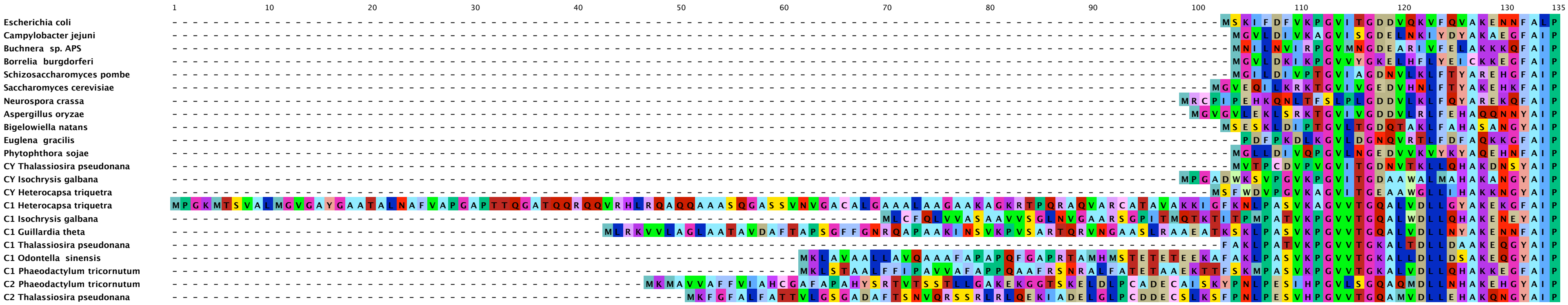
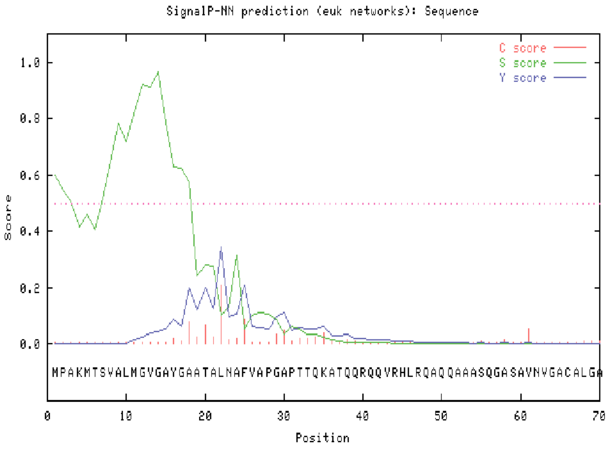
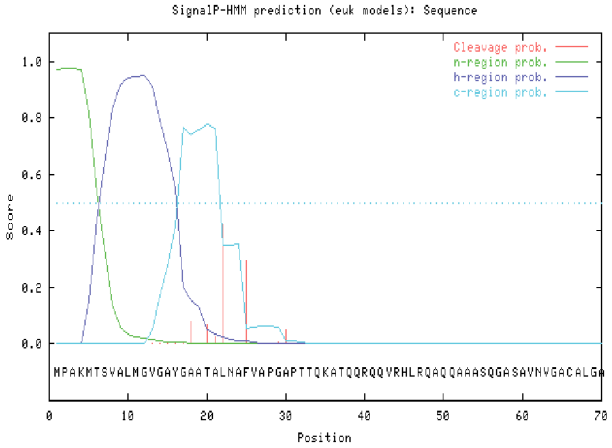


```
                              1        10        20        30        40        50        60        70        80        90        100       110       120       130  135
Escherichia coli              ------------------------------------------------------------------------------------------------------MSKIFDFVKPGVITGDDVQKVFQVAKENNFALP
Campylobacter jejuni          -------------------------------------------------------------------------------------------------------MGVLDIVKAGVISGDELNKIYDYAKAEGFAIP
Buchnera sp. APS              -------------------------------------------------------------------------------------------------------MNILNVIRPGVMNGDEARIVFELAKKKQFAIP
Borrelia burgdorferi          -------------------------------------------------------------------------------------------------------MGVLDKIKPGVVYGKELHFLYEICKKEGFAIP
Schizosaccharomyces pombe     -------------------------------------------------------------------------------------------------------MGILDIVPTGVIAGDNVLKLFTYAREHGFAIP
Saccharomyces cerevisiae      -----------------------------------------------------------------------------------------------------MGVEQILKRKTGVIVGEDVHNLFTYAKEHKFAIP
Neurospora crassa             --------------------------------------------------------------------------------------------------MRCPIPEHKQNLTFSLPLGDDVLKLFQYAREKQFAIP
Aspergillus oryzae            ---------------------------------------------------------------------------------------------------MGVGVLEKLSRKTGVIVGDDVLRLFEHAQQNNYAIP
Bigelowiella natans           ------------------------------------------------------------------------------------------------------MSESKLDIPTGVLTGDQTAKLFAHASANGYAIP
Euglena gracilis              --------------------------------------------------------------------------------------------------------PDFPKDLKGVLDGNQVRTLFDFAQKKGFAIP
Phytophthora sojae            -------------------------------------------------------------------------------------------------------MGLLDIVQPGVLNGEDVVKVYKYAQEHNFAIP
CY Thalassiosira pseudonana   -------------------------------------------------------------------------------------------------------MVTPCDVPVGVITGDNVTKLLQHAKDNSYAIP
CY Isochrysis galbana         --------------------------------------------------------------------------------------------------MPGADWKSVPGVKPGVITGDAAWALMAHAKANGYAIP
CY Heterocapsa triquetra      -----------------------------------------------------------------------------------------------------MSFWDVPGVKAGVITGEAAWGLLIHAKKNGYAIP
C1 Heterocapsa triquetra      MPGKMTSVALMGVGAYGAATALNAFVAPGAPTTQGATQQRQQVRHLRQAQQAAASQGASSVNVGACALGAAALAAGAAKAGKRTPQRAQVARCATAVAKKIGFKNLPASVKAGVVTGQALVDLLGYAKEKGFAIP
C1 Isochrysis galbana         ---------------------------------------------------------------------MLCFQLVVASAAVVSGLNVGAARSGPITMQTKTITPMPATVKPGVVTGQALWDLLQHAKENEYAIP
C1 Guillardia theta           ------------------------------------------MLRKVVLAGLAATAVDAFTAPSGFFGNRQAPAAKINSVKPVSARTQRVNGAASLRAAEATKSKLPASVKPGVVTGQALVDLLNYAKENNFAIP
C1 Thalassiosira pseudonana   ------------------------------------------------------------------------------------------------------FAKLPATVKPGVVTGKALTDLLDAAKEQGYAIP
C1 Odontella sinensis         -------------------------------------------------------------MKLAVAALLAVQAAAFAPAPQFGAPRTAMHMSTETETEEKAFAKLPASVKPGVVTGKALLDLLDSAKEQGYAIP
C1 Phaeodactylum tricornutum  -------------------------------------------------------------MKLSTAALFFIPAVVAFAPPQAAFRSNRALFATETAAEKTTFSKMPASVKPGVVTGKALVDLLQHAKKEGFAIP
C2 Phaeodactylum tricornutum  ----------------------------------------------MKMAVVAFFVIAHCGAFAPAHYSRTVTSSTLLGAKEKGGTSKELDLPCADECAISKYPNLPESIHPGVLSGQAQMDLLNHAKEHGYAIP
C2 Thalassiosira pseudonana   --------------------------------------------------MKFGFALFATTVLGSGADAFTSNVQRSSRLRLQEKIADELGLPCDDECSLKSFPNLPESVHPGVVTGQAMVDLLEHAKQNGYAIP
```

Figure S2A: *Heterocapsa triquetra* C1

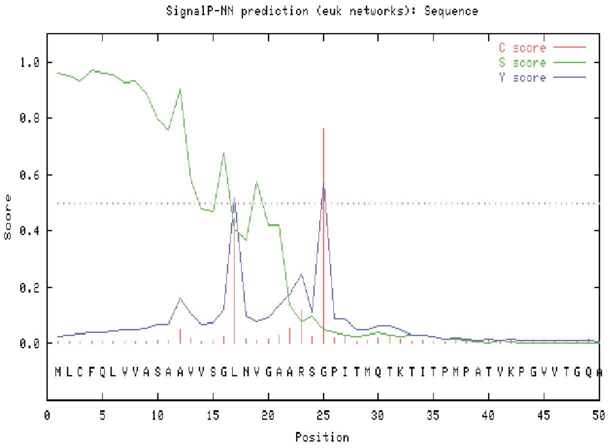


```
# Measure  Position  Value  Cutoff  signal peptide?
  max. C    22       0.209   0.32   NO
  max. Y    22       0.344   0.33   YES
  max. S    14       0.964   0.87   YES
  mean S     1-21    0.602   0.48   YES
       D     1-21    0.473   0.43   YES
# Most likely cleavage site between pos. 21 and 22: ATA-LN
```

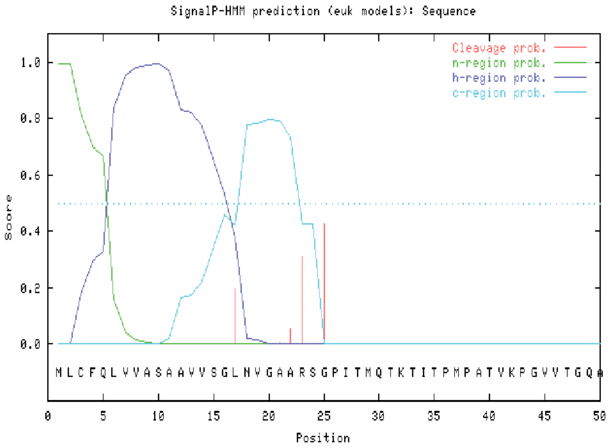


Prediction: Signal peptide
Signal peptide probability: 0.969
Signal anchor probability: 0.005
Max cleavage site probability: 0.423 between pos. 21 and 22

Figure S2B: *Isochrysis galbana* C1

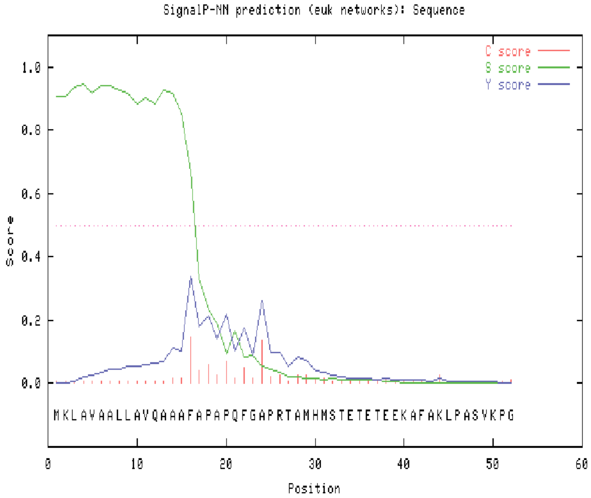


```
# Measure  Position  Value  Cutoff  signal peptide?
  max. C    25       0.761   0.32   YES
  max. Y    25       0.569   0.33   YES
  max. S    4        0.968   0.87   YES
  mean S     1-24    0.652   0.48   YES
       D     1-24    0.610   0.43   YES
# Most likely cleavage site between pos. 24 and 25: ARS-GP
```

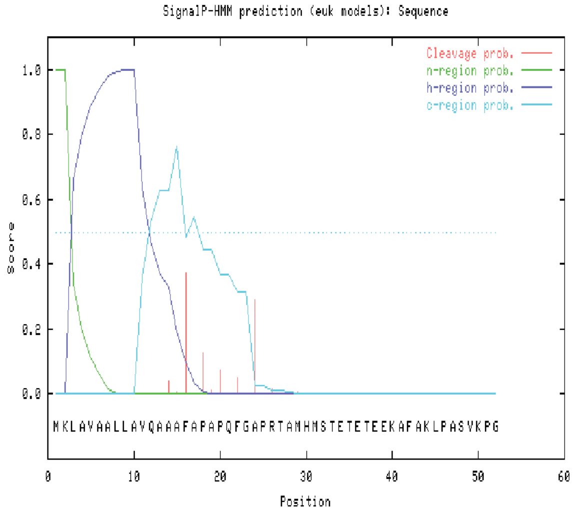


Prediction: Signal peptide
Signal peptide probability: 0.996
Signal anchor probability: 0.000
Max cleavage site probability: 0.424 between pos. 24 and 25

Figure S2C: *Odontella sinensis* C1

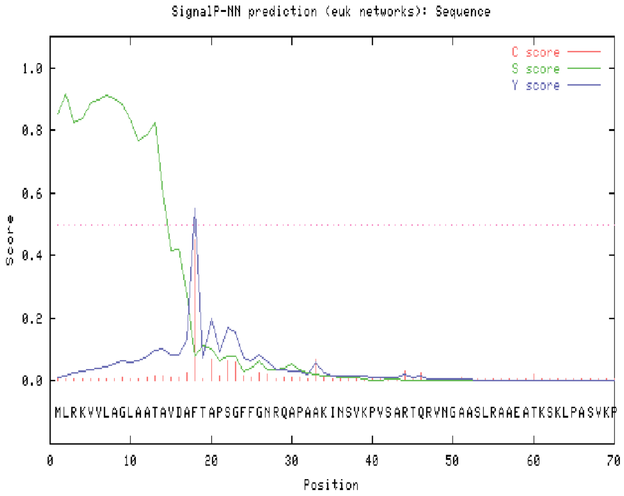


```
# Measure  Position  Value  Cutoff  signal peptide?
  max. C    16      0.148   0.32   NO
  max. Y    16      0.337   0.33   YES
  max. S     4      0.944   0.87   YES
  mean S    1-15    0.914   0.48   YES
       D    1-15    0.625   0.43   YES
# Most likely cleavage site between pos. 15 and 16: AAA-FA
```

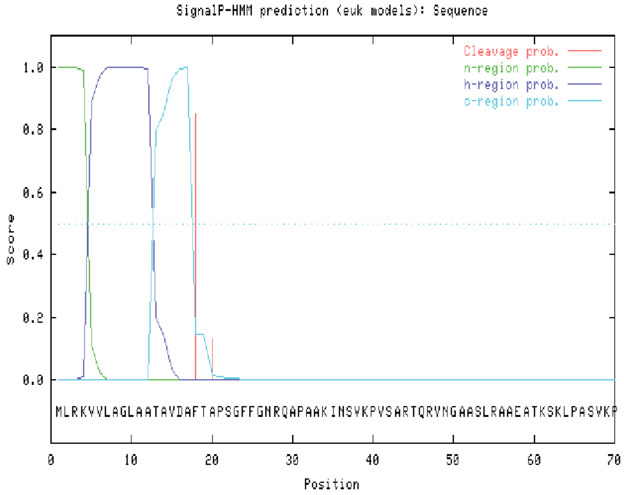


```
Prediction: Signal peptide
Signal peptide probability: 1.000
Signal anchor probability: 0.000
Max cleavage site probability: 0.373 between pos. 15 and 16
```

Figure S2D: *Guillardia theta* C1

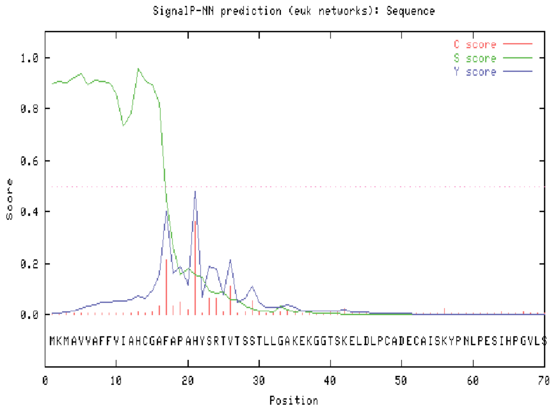


```
# Measure  Position  Value  Cutoff  signal peptide?
  max. C    18      0.452   0.32   YES
  max. Y    18      0.551   0.33   YES
  max. S     2      0.915   0.87   YES
  mean S    1-17    0.756   0.48   YES
       D    1-17    0.653   0.43   YES
# Most likely cleavage site between pos. 17 and 18: VDA-FT
```

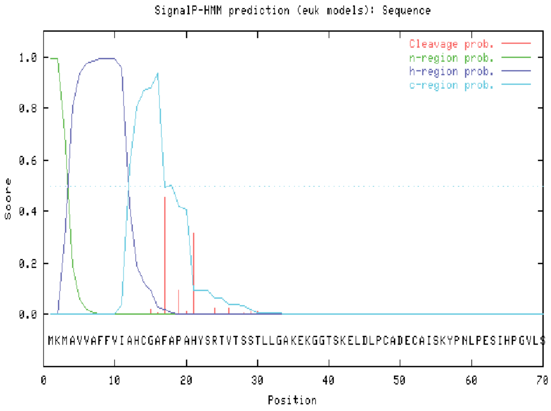


```
Prediction: Signal peptide
Signal peptide probability: 0.998
Signal anchor probability: 0.000
Max cleavage site probability: 0.852 between pos. 17 and 18
```

Figure S3A: *Phaeodactylum tricornutum*C2

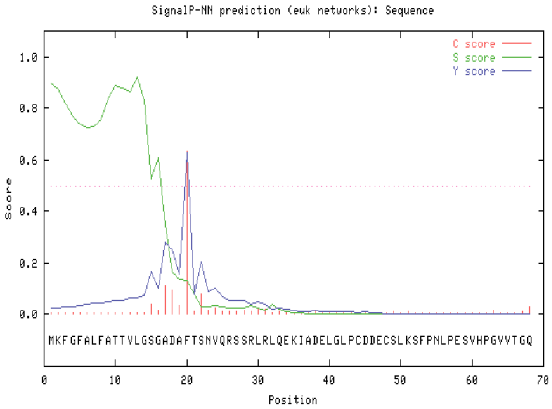


```
# Measure  Position  Value  Cutoff  signal peptide?
 max. C    21      0.365   0.32   YES
 max. Y    21      0.480   0.33   YES
 max. S    13      0.954   0.87   YES
 mean S     1-20    0.759   0.48   YES
     D     1-20    0.620   0.43   YES
# Most likely cleavage site between pos. 20 and 21: APA-HY
```

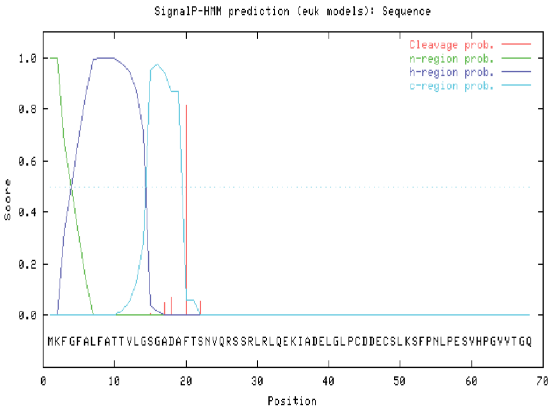


```
Prediction: Signal peptide
Signal peptide probability: 0.995
Signal anchor probability: 0.000
Max cleavage site probability: 0.456 between pos. 16 and 17
```

Figure S3B: *Thalassiosira psuedonana* C2

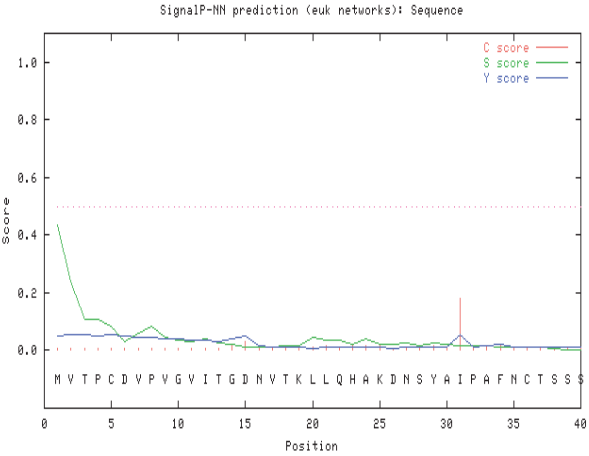


```
# Measure  Position  Value  Cutoff  signal peptide?
 max. C    20      0.633   0.32   YES
 max. Y    20      0.626   0.33   YES
 max. S    13      0.921   0.87   YES
 mean S     1-19    0.701   0.48   YES
     D     1-19    0.664   0.43   YES
# Most likely cleavage site between pos. 19 and 20: ADA-FT
```

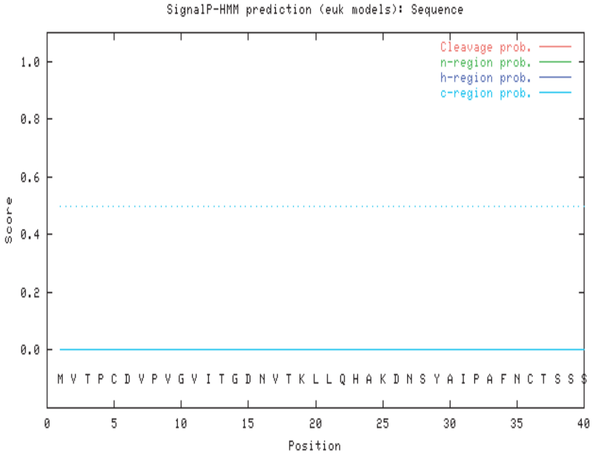


```
Prediction: Signal peptide
Signal peptide probability: 0.998
Signal anchor probability: 0.000
Max cleavage site probability: 0.814 between pos. 19 and 20
```

Figure S4A: *Thalassiosira pseudonana* CY

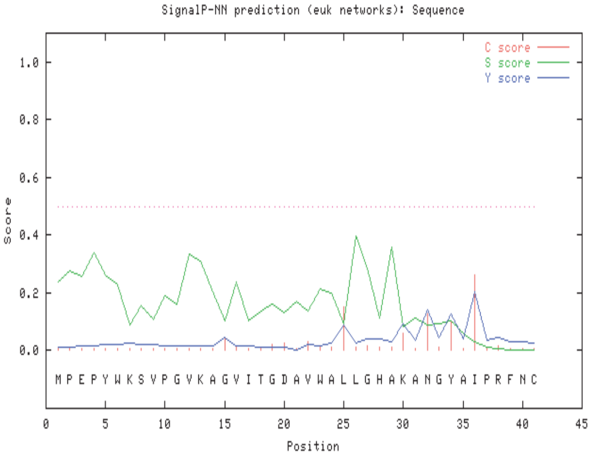


```
# Measure  Position  Value  Cutoff  signal peptide?
  max. C    31       0.180   0.32   NO
  max. Y    31       0.055   0.33   NO
  max. S     1       0.437   0.87   NO
  mean S    1-30     0.058   0.48   NO
       D    1-30     0.056   0.43   NO
```

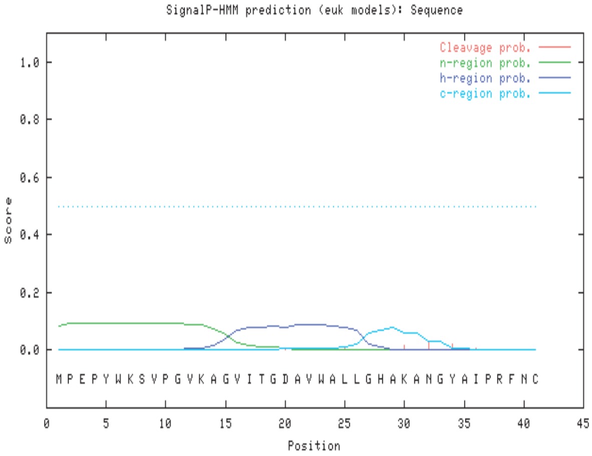


Prediction: Non-secretory protein
Signal peptide probability: 0.000
Signal anchor probability: 0.000

Figure S4B: *Isochrysis galbana* CY

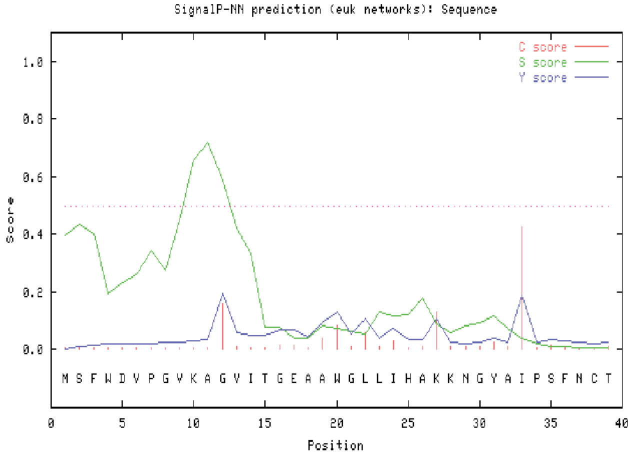


```
# Measure  Position  Value  Cutoff  signal peptide?
  max. C    36       0.262   0.32   NO
  max. Y    36       0.206   0.33   NO
  max. S    26       0.398   0.87   NO
  mean S    1-35     0.187   0.48   NO
       D    1-35     0.197   0.43   NO
```

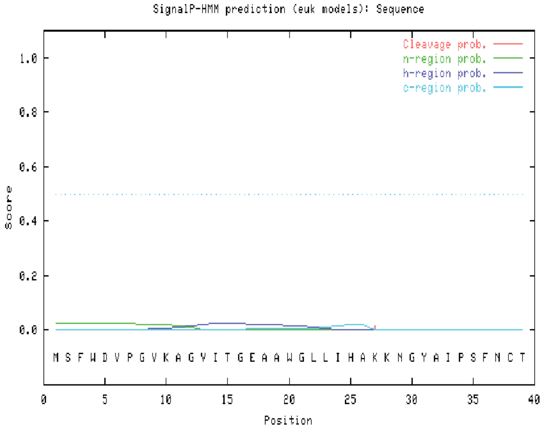


Prediction: Non-secretory protein
Signal peptide probability: 0.084
Signal anchor probability: 0.011

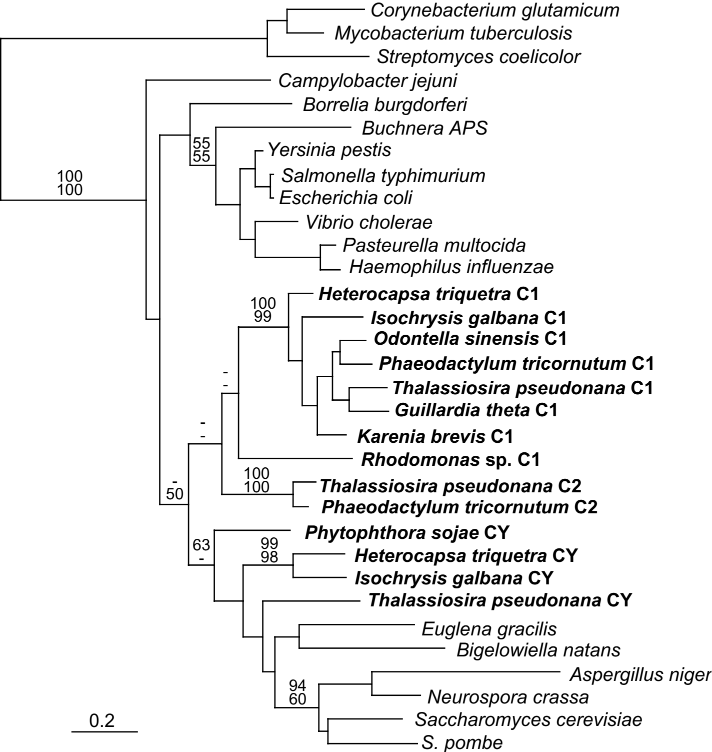
Figure S4C: *Heterocapsa triquetra* CY



```
# Measure  Position  Value  Cutoff  signal peptide?
  max. C    33       0.428   0.32   YES
  max. Y    12       0.196   0.33   NO
  max. S    11       0.718   0.87   NO
  mean S     1-11    0.397   0.48   NO
       D     1-11    0.297   0.43   NO
```



```
Prediction: Non-secretory protein
Signal peptide probability: 0.027
Signal anchor probability: 0.000
Max cleavage site probability: 0.017 between pos. 26 and 27
```


Corynebacterium glutamicum
Mycobacterium tuberculosis
Streptomyces coelicolor
Campylobacter jejuni
Borrelia burgdorferi
Buchnera APS
55
55
Yersinia pestis
100
100
Salmonella typhimurium
Escherichia coli
Vibrio cholerae
Pasteurella multocida
Haemophilus influenzae
Heterocapsa triquetra C1
100
99
Isochrysis galbana C1
Odontella sinensis C1
Phaeodactylum tricornutum C1
-
-
Thalassiosira pseudonana C1
Guillardia theta C1
-
-
Karenia brevis C1
Rhodomonas sp. C1
100
100
-
50
Thalassiosira pseudonana C2
Phaeodactylum tricornutum C2
Phytophthora sojae CY
63
-
99
98
Heterocapsa triquetra CY
Isochrysis galbana CY
Thalassiosira pseudonana CY
Euglena gracilis
Bigelowiella natans
Aspergillus niger
94
60
Neurospora crassa
Saccharomyces cerevisiae
0.2
S. pombe

Supplementary Figure Legends.

Figure S1. Alignment of the amino terminus of representative class II type A FBA genes, including all chromalveolate sequences. Chromalveolates are designated CY (cytosolic), C1 (plastid-targeted) and C2 (diatom specific plastid-targeted). The *Thalassiosira* C1 sequence is truncated.

Figure S2. Signal peptide predictions for FBA C1 sequences from (A) *Heterocapsa*, (B) *Isochrysis*, (C) *Odontella*, and (D) *Guillardia*. In each part, the top chart is the neural network prediction and the lower chart is the hidden Markov model (HMM) prediction. All sequences are predicted to encode signal peptides and both models predict similar or identical cut sites. All predictions carried out using SignalP version 3.0 (1).

Figure S3. Signal peptide predictions for FBA C2 sequences from (A) *Phaeodactylum*, and (B) *Thalassiosira*. Both sequences are predicted to encode signal peptides and both models predict similar or identical cut sites. Descriptions are as in Figure S2.

Figure S4. Signal peptide predictions for FBA CY sequences from (A) *Thalassiosira*, (B) *Isochrysis*, and (C) *Heterocapsa*. Descriptions are as in Figure S2.

Figure S5. Protein maximum likelihood phylogeny of class II type A FBA including partial sequences from the cryptomonad *Rhodomonas* and the haptophyte plastid-containing dinoflagellate *Karenia*. The phylogeny is inferred using PhyML and numbers at nodes correspond to PhyML bootstrap support. *Rhodomonas* (the much shorter sequence) is sister to the C1 clade and *Karenia* falls within C1 with strong support. *Karenia* shows no affinity to either *Isochrysis* or *Heterocapsa*, so the source of this enzyme is not clear.

1. **Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak.** 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. **340:**783-795.