

Supplementary Information

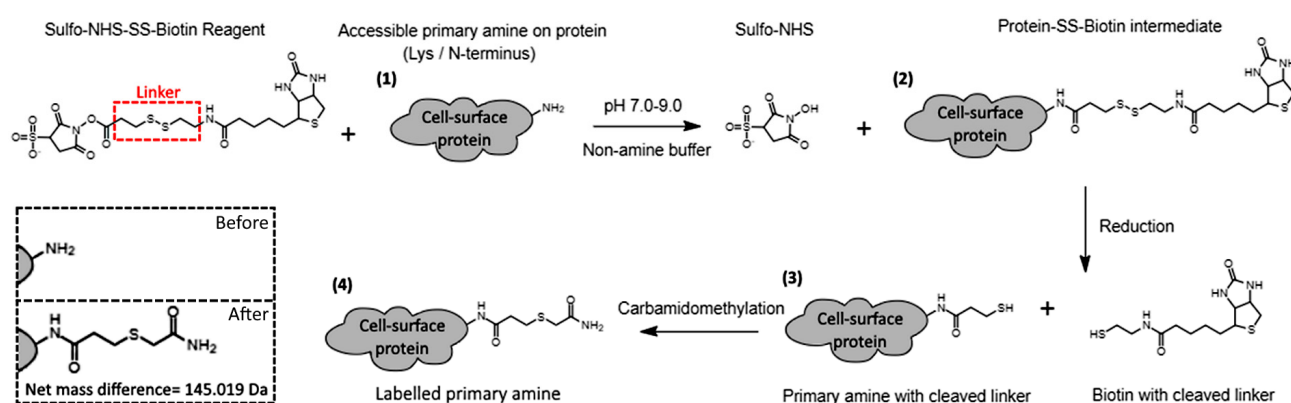
# Finding Species-Specific Extracellular Surface-Facing Proteomes in Toxic Dinoflagellates

Kenrick Kai-yuen Chan <sup>1</sup>, Hang-kin Kong <sup>1,2</sup>, Sirius Pui-kam Tse <sup>1</sup>, Zoe Chan <sup>1</sup>, Pak-yeung Lo <sup>1</sup>, Kevin W.H. Kwok <sup>1,2</sup> and Samuel Chun-lap Lo <sup>1,2</sup> \*

<sup>1</sup> Department of Applied Biology and Chemical Technology, Faculty of Applied Science and Textiles, The Hong Kong Polytechnic University, Hung Hom, Hong Kong; kai-yuen.ky.chan@connect.polyu.hk (K.K.C.); hang-kin.kong@polyu.edu.hk (H.K.); pktse@polyu.edu.hk (S.P.T.); zoe-z.c.chan@polyu.edu.hk (Z.C.); pypatlo@polyu.edu.hk (P.L.); kwh.kwok@polyu.edu.hk (K.W.H.K.)

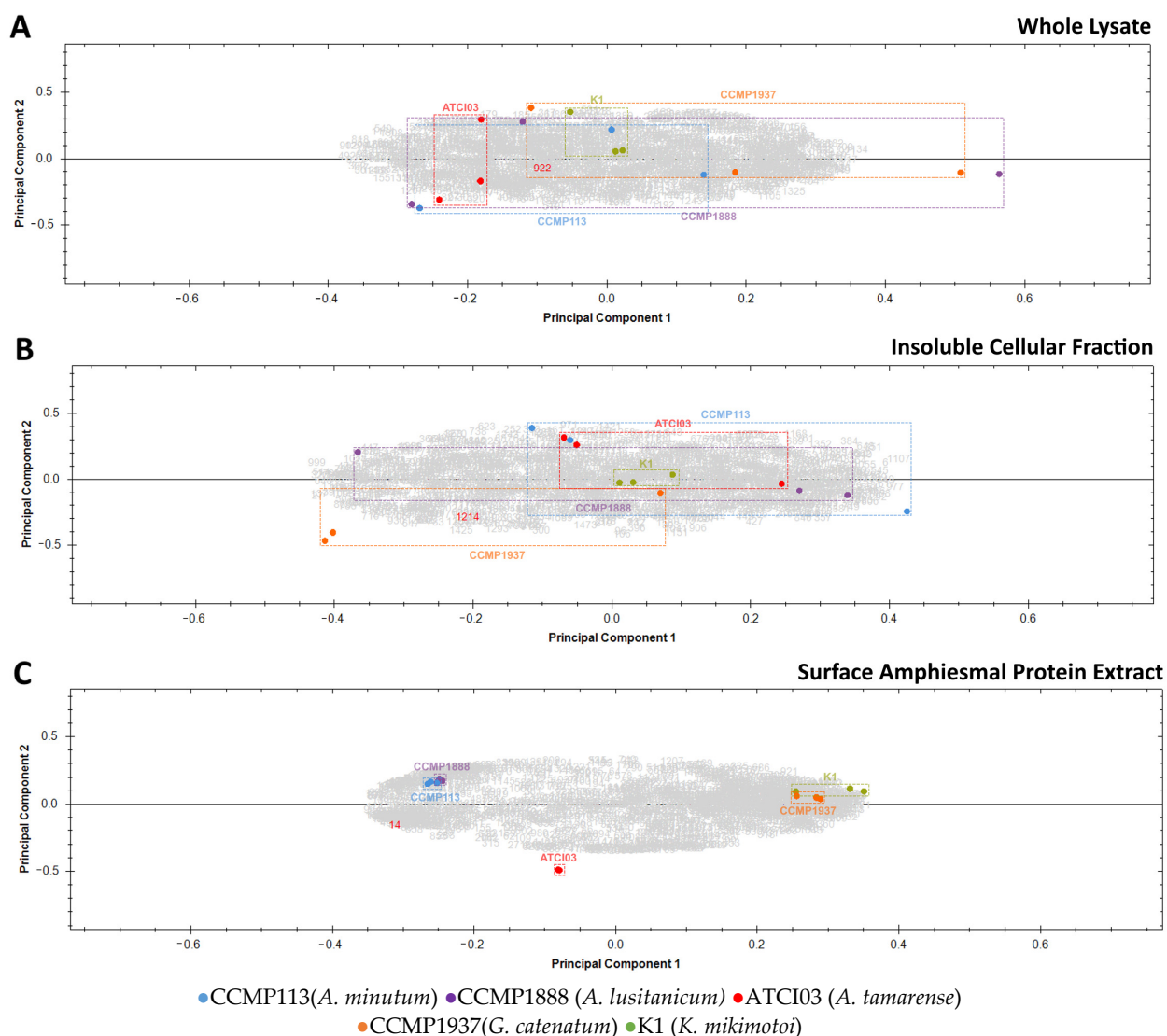
<sup>2</sup> Research Institute for Future Food, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

\* Correspondence: Samuel.chun-lap.lo@polyu.edu.hk; Tel.: +852-34008669; Fax: +852-23649932



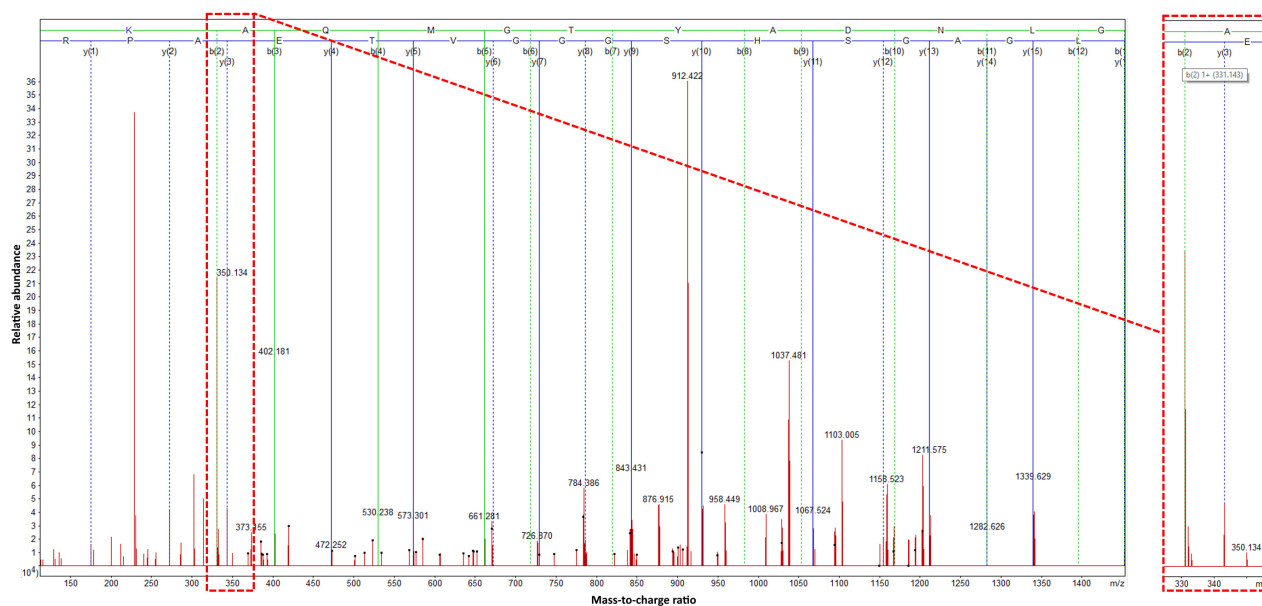
**Figure S1.** Reaction scheme of surface labeling using sulfo-NHS-SS-biotin reagent.

Four major reactions involved: (1) The exposed primary amines on the cell surface proteins reacted with the *N*-Hydroxy-succinimide (NHS) of the labeling reagent and yielded protein-SS-biotin intermediates. (2) The disulfide bond on the intermediate was reduced and disconnected by dithiothreitol (DTT) reduction during the mass spectrometry sample preparation procedure, and thus the side chain with biotin tag was removed. Primary amine was labeled partially with the leftover linker. Biotin with partial linker was removed during the C18 reverse phase chromatography. (3) The thiol group of the partially labelled primary amine was carbamidomethylated with iodoacetamide (IAA). (4) The completely labelled primary amine on a peptide would provide an increased mass difference of 145.019 Da.



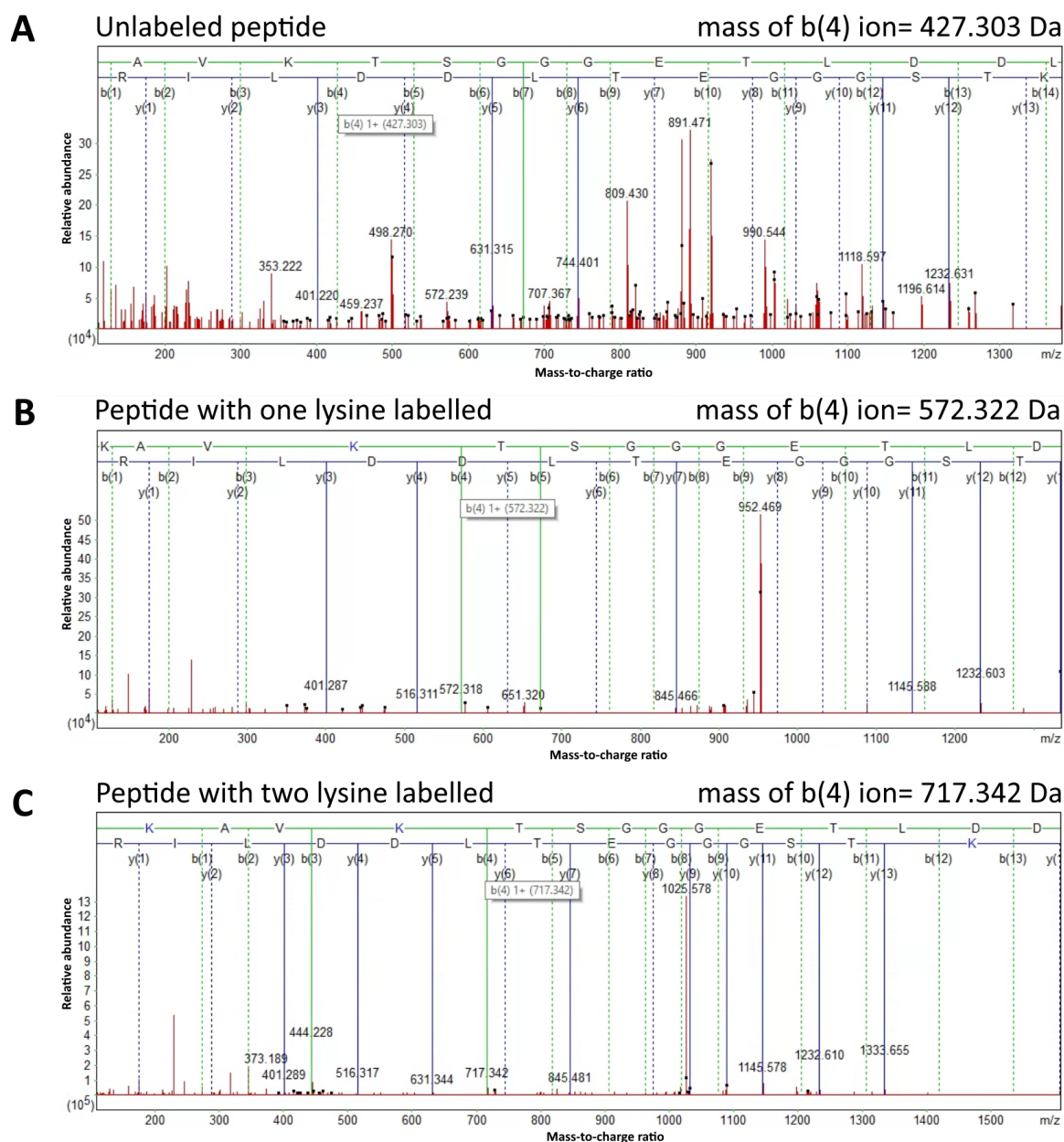
**Figure S2.** Difference in principal components of (A) whole lysates, (B) insoluble cellular fractions and (C) surface amphiesmal protein extracts of the five species studied.

Area of the boxes with dashed line indicated the consistence of the samples from same species. The larger of the box represented less consistence of the samples from the same species. Among three methods of extraction, novel amphiesmal extracts showed the best consistence and species coherence comparing to whole lysates and insoluble cellular fractions.



**Figure S3a.** MS/MS spectrum of a labeled peptide from a specific ortholog of K1.

The amino acid sequence of the peptide identified was “GKAQMGTYADNLGAGSHSGGGVTEAPR” and the actual mass of its b(2) ion was 331.143 Da. The tagged peptide was accurately identified by the LC-Orbitrap-MS that the difference between the actual and expected mass of the b(2) ion of labeled peptide was 0.00025 Da, which was within the mass tolerance and thus neglectable. Also, the modification was occurred on the expected primary amine containing lysine residue. This validated the applicability of this labeling reagent towards dinoflagellates. More importantly, mutual authentication from the annotation databases and the tag significantly indicated the confidence of the identified cell-surface proteins.



**Figure S3b.** Mass difference of the same peptide species identified in K1 with different numbers of lysine residue labelled: (A) Unlabeled peptide; (B) Peptide with one lysine residue labeled; (C) Peptide with two lysine residues labeled.

Another example of a peptide species identified in K1 with different numbers of labeled lysine residue. The amino acid sequence of the peptide species identified was “KAVKTSGGGETLDDLIR”. This peptide contained two possible residues (lysine) being labelled. In the native unlabeled peptide, the actual mass of the b(4) ion was 427.303 Da. Increases of 145.019 Da and 290.039 Da were observed in the b(4) ion of the singly labeled peptide shown in (B) and doubly labeled peptide shown in (C) respectively.

Table S1. DNA sequences of the PCR products.

ITS-DNA Sequence(s) of CCMP113
<p>&gt;CCMP113_Pair1</p> <p>TCTGCTCCACGCGTTGGGAGCTCTCCCATATGGTTCGACCTGCAGGCGGCGCGAATTCAGTAGTGATTTCCTCCGCTTATTGATATGCTTAAATTCAGCAGGTTTACATGCTTCACTTCATGTTAAGTTTACGCTGTTTCAGCTACCCAGACAAATACTATGCAACCAATGCAAGTGACAAAGGCTAGACCTAAGCAAACACAATGAGGTTGCAATGCATTTCGCAAGCGCGGGTGTCGAATGCAACAGCATTGACACACACAGCTCACAACGTTGCTGAACAGCAAATTCAATGTTGGAAGTTAAGTGACATTGAATCAAGCACACCTTCAAGCATATCCCGAAGGTGC AAATTACGTTCAAACATCTATTGGCTCACGGAATTCTGCAATTCACAATGCATATCACATTTTGCTGCATTCTTCATCATCAATTGA GCTAAGACATTCAATTGCAAAAACACATTTACATTTCAAGTTCAATGCAAGAAAACAAGTACAAGCAATCAACAAAAGGTCATCAACAC AGCAGCATGAAAGCGCTTGCAGAAGCAAGCCATGCCACCCACGAGCAAATGATGCAGAACCTAAAGCCTAGGAACCCACACAGCT CACAAAGTCGTGAAAGATTAATGTTGTTAGGAATGTGCAAATGATCCTTCCGCGAGGTTACCTACGGAAATCGAATTCGCCGCGCCG CCATGGCGGCGGGAGCATGCGACGTCGGGCCCAATTGCGCCTTAGTGAGTCGTATTACAATTCACTGGCCGTGCTTTTACAAAGT CGTGACTGGGAAAACCTGGCGTTACCCAACCTTAATCGCCTTGACGACACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCC CGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGG TGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCTTTTCGCTTTCTTCCCTTCTTTCTCGCACGTTTCGCG CTTTCCCGCTCAGCTCTAATCGGGGCTCCTTAGGTTGATTTAGTGCTTACGCACTCGACCCCAAACTTGATAGTGATGATCACGTA ATGGCAATCGCCTGATAGACGATTTTTCGCTGACTGGTCCACGTCCTTAATATGACTCTTGTTCACACCTGG</p> <p>&gt;CCMP113_Pair2</p> <p>ACCAAGTCGCATGCTCCCGGCCGCCATGGCGGCCGCGGAATTCGATTAGAACCTTACCCTTAGAGGAAGGAGAAGTCGTAAACAAG GTTTCTGTAGGTGAACCTGCGGAAGGATCATTTGCACATCCCTAACAACATTAATCTTTCACGACTTTGTGAGCTGTGGTGGGGTTC CTAGGCTTTAGGTTCTGCATCATTTGCTCGTGGGTGGCATGGCTTGCTTCTGCAAGCGCTTTCATGCTGCTGTGTTGATGACCTTTT GTTGATTGCTTGTACTTGTCTTTCGATTGAACCTGAATGTGAAATGTGTTTTGCAATGAATGTCTTAGCTCAATTGATGATGAAG AATGCAGCAAAATGTGATATGCATTGTGAATTGCAGAATTCGGTGAGCCAATAGATGTTTGAACGTAATTTGCACCTTCGGGATATG CTTGAAGGTGTGCTTGATTCAATGTCAATTAACCTTCCAACATTGAATTTGCTGTTTCAGCAACGTTGTGAGCTGTGTGTGCAATGCT GTTGCAATTGGACACCCGCGCTTGCGAATGCATTGCAACCTCATTTGTGTTTGTCTTAGGTCTAGCCTTTGTCACTTGCATTGGTTGCAT AGTATTTGTCTGGGTAGCTGAACAGCGTAACTTAACATGAAGTGAAGCATGTAAACCTGCTGAATTTAAGCATATAAGTAGGCGGT GGAAATGAACCAAATGGGATTCTTAAAGTAATTGCGAATGAAGAGGGATTTGCTTAGCAATCACTAGTGAATTCGCGGCCGCCCTGC AGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCAT GGTCAAGTCTGTTTCTGTGTGAAATTGTTATCCGCTCACATTTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGG GTGCCTAATGAGTGAGCTAATCACTCAATTAATTGCGTTGCGCTCACTGCCCGCTTTCAGTCCGGAAACCTGTCTGTCGCGAGCTGCATT AATGAATCGGCCACGCGCGGGGAGGAGCGGTTTTCGCTATTGGGCGCTCTTTCGCTTTCTCGCTCACTGACTCGCTGCGCTCGTCTTC GCTGCGGCGAGGCGTATCAGCTCACTCAAAGGCGGTATTTCGCTTATCCAAGGATTTCAGGGATACGCAGGAAGACTGGTGGAGCAGGC CAGCTAGCTCAGGACTCGTAAAAGGGCCCCGCTGTGTCTGTTGCGT</p>
ITS-DNA Sequence(s) of CCMP1888
<p>&gt;CCMP1888_Pair1</p> <p>CTCAGCTCCCGGCCGCCATGGCGGCCGCGGAATTCGATTTCCTCCGCTTATTGATATGCTTAAATTCAGCAGGTTTACATGCTTCA CTTTCATGTTAAGTTTACGCTGTTTCAGCTACCCAGACAAATACTATGCAACCAATGCAAGTGACAAAGGCTAGACCTAAGCAAACACA ATGAGGTTGCAATGCATTTCGCAAGCGCGGGTGTCGAATGCAACAGCATTGACACACACAGCTCACAACGTTGCTGAACAGCAAATTC AATGTTGGAAGTTAATTGACATTGAATCAAGCACACCTTCAAGCATATCCCGAAGGTGCAAATTACGTTCAAACATCTATTGGCTCA CGGAATTCTGCAATTCAATGCATATCACATTTTGCTGCATTCTTCATCATCAATTGAGCTAAGACATTCAATTGCAAAAACACATT TCACATTCAAGTTCAATGCAAGAAAACAAGTACAAGCAATCAACAAAAGGTCATCAACACAGCAGCATGAAAGCGCTTGCAGAAGCAA GCCATGCCACCCACGAGCAAATGATGCAGAACCTAAAGCCTAGGAACCCACACAGCTCACAAGTTCGTGAAAGATTAATGTTGTT AGGAATGTGCAAATGATCCTTCCGCGAGGTTACCTACGGAAATCACTAGTGAATTCGCGGCCGCCCTGCAGGTCGACCATATGGGAGA GCTCCCAACGCGTTGGATGCATAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGT GTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTA ACTCACATTAATTGCGTTGCGCTCACTGCCCCGCTTTCCAGTCCGGAAACCTGTGCTGCCAGCTGCATTAAATGAATCGGCCAACGCGC GGGGAGAGGCGGTTTTCGCTATTGGGCGCTCTTTCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTCGCTCGGTCGCGAGCGGTATC AGCTCACTCAAGCGTATACGTTATCCACAGAATCAGGGATACGCAGAAAGAACATGGTGGAGCAAAGGCAGCAAGTAGATCGTAAAA GGCCCGGATGCTGGCGTTTCCCTAGGTTTCGCTCCTTGACCAGACTTCGAATTCGCACGCTCAAGTCCCTAAGGGTGGTGC</p>
ITS-DNA Sequence(s) of ATCI03
<p>&gt;ATCI03_Pair2</p> <p>AGTCATCCAACGCGTTGGGAGCTCTCCCATATGGTTCGACCTGCAGGCGGCGCGAATTCAGTAGTGATTGCTAAGCAAATCCCTCTT CATTTCGCAATTAAGTGAAGAGATCCAGTTGGTTTAAATTTCCACCCTTACTTATATGCTTAAATTCAGCAGGTTTATGTGCTTCACT TCACGCTTAGCATATGCAGGTTCTTCAGCTTGTCCAGTTGTGCACATTTACAAATATCAACCATCAAATGACGAAACAGGTCTGCCC CACGAAACCATGTTTAGGTTGTAAACAACCTGTATGACATCCAGGTTCAATGCAACACATTGACCTTTATTGTTTCAACAATGCTAA TCAGCATTAAGTGTATATGAAAGCTACTTTGCATTAAACCAAGCAAACCTTCAAGAATATCCCAAAGGTACAAGTAACATTCAAACA TCTGTTAGCTCACGGAATTCTGCAATTGCAATGCATATTGCATGTTGCTGCATTCTTCATCATTTATTGAGCTAAGACATTCCTTG CAAAACATAAGTGTGTCAGCAACTGGAGAAATGCAACAGTTTACACAACACATCACACGCTACAGCACACATGCTTGAAAAACAAA TCATGCTGCCAGGGCAAGTGCAAGTCCATTGCAATACATGCTTACAGTTGCCACAATATCATTAGCATGCAGTTGGCTACACATG TGCAATGATCCTTCTGCAGGTTACCTACAGAAACCTTGTACGACTTCTCTTCTCTAAGTGGTAAGGTTCAAATCGAATTCCT CCGGCCGCCATGGCGGCCGGGAGCATGCGACGTCGGGCCAATTCGCCCTTAGTGAGTCGTATTACAATTCACTGGCCGTGCTTTT</p>

ACAACGTCGTGACTGGGAAAACCCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTTCGCCAGCTGGCGTAATAGCGA  
AGAGGCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGG  
GTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCA

### ITS-DNA Sequence(s) of CCMP1937

>CCMP1937\_Pair1

AAAGGCCATCCAACGCGTTGGGAGCTCTCCCATATGGTTCGACCTGCAGGCGGCCGCGAATTCACTAGTGATTTCTCCGCTTATTGA  
TATGCTTAAATTCAGCGGGTTTGCCAACCTAATTCATGTCTTAAAGCTGGGCGCAGCCACAGTTGATCGATGCGAATGAAACATTG  
ACACGTCGTGCGCAACTAGTTGATTTCGCTTCAATCAGTTGGCAGCGTGACAACACCTTGTCAATCGAATGCCGCAAGAGCGCCACAC  
ACCCTCTCGTCAGAGGACGCGCGTCTTGAAGCAGCATAATCACAGCAGAACGCACCTTCGAGAGGAAGAATGACAACCTGCTGGAAA  
TAACAACATAGACACTGAAGCAGGCATGCTTTAGGGATATCCAAAAGCACAATATGCGTTCAAGTTTCTGTGCGTTTCACGGAATT  
CTGCAATTCACAATGCTTATCGCACTTCGCTGCGCCCTTCATCGTTGTTTGAACCGAGACATCCGTCGCTGAAAGTTGTAATTGCTG  
GTTGATACCAATGCAACGAATACAGCTTCACATTAGCCTGAGAGACTTCAGGTAGCAGGCACAAGGCGATTAAAGCACCGCCTTGCC  
CCTGCTACCCCTCGAGATTGTTTCGTGGCGGCAACTAGAGGAGCAACTTCACGTCACAGCCTCACAGAAGAGTTTACAAGAATGGTTTT  
GGATACGTGCGAATGATCCTTCCGCGAGTTTACCTACGGAAAATCGAATTTCCGCGGCCCGCCATGGCGGCCGGGAGCATGCGACGTCG  
GGCCCAATTGCGCCTATAGTGAGTCGTATTACAATTCAGTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCCTGGCGTTACCC  
AACTTAATCGCCTTGCAGCACATCCCCCTTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGC  
GCAGCCTGAATGGCGAATGGACGCGCCCTGTAGCGCGCATTAAGCGCGGCGGGTGTGTGGTTACGCGCAGCGTGACCGCTACACTT  
GCCAGCGCCCTAGCGCCGCTCCTTTTCGCTTTCTTC

### ITS-DNA Sequence(s) of K1

>K1\_Pair1

AAGCTCCACGCGTTGGGAGCTCTCCCATATGGTTCGACCTGCAGGCGGCCGCGAATTCACTAGTGATTTGAACCTTATCACTTAGAGG  
AAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTACACGCATCCAACCTGAATTCATTGTGAACCATCT  
TCTGTGAGTTGTGCTCTTGGCTGGAGACAACCTTGTCTACAACCTTGTGCGAATCTCGCGGGTAGTTTCGGGTGATATTGGACCGTCT  
TCCTCTATTGTCTGGCTACCATTTGTTGTCTGACCAATGTCTTGCATGATGACCTATCTTTTAAAGATGTGGGTGACCACATGTTTGA  
TTCACATACAACCTTTTCAGCGACGGATGTCTCGGCTCGAACAACGATGAAGGGCGCAGCGAAGTGTGATAAGCATTTGTGAATTGCAGA  
ATTCCGTGAACCAATAGGGATTTGAACGTATACTGCGCTTTTCGGGATATCCCTGAAAGCATGCCTGCTTCAGTGTCAATATTTCTCTC  
ATGCCACTGTCTCTTGTCTGTGCTTTGCACATGCAACTGACAGCAGTGTGTGCTGTGTCATTAAAGTGCTCTTCTGCCCCGTGAT  
GCATTGAATCCATGGGTTTTGTGCTCTGTCTGCACTTCAATTTGTGGCTTCTTCCATGACATGAAGTTAGGTAAAGCAAACCCGCTGAATTTAAGCA  
TATAAGTAAGCGGAGGATAAGAACTAAATAGGATTCCCTCAGTAATGGCGAATGAACAGGGATTAGCTCAGCAATCGAATTTCCCGC  
GGCCGCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTGCGCCTATAGTGAGTCGTATTACATTTCACTGGCCGTCGTTTTAC  
ACGTCGTGACTGGGAAAACCCCTGGCGTTACCCAACCTTAATCGCCTTGCAGCACATCCCCCTTTTCGCCAGCTGGCGTAATAGCGAAGA  
AGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGGCGAATGGGAACGCGGCCCTGTTAACGGGCCGCCATTATAC  
CCCGCCGGGGGGGTGGTGGGG

>K1\_Pair2

TTTCTCCAACGCGTTGGGAAGCTCTCCCATATGGTTCGACCTGCAGGCGGCCGCGAATTCACTAGTGATTTCCGTAGGTGAACCTGCG  
GAAGGATCATTACACGCATCCAACCTGAATTCATTGTGAACCATCTTCTGTGAGTTGTGCTCTTGGCTGGAGACAACCTTGTCTACA  
ACCTTGTGCGAATCTCGCGGGTAGTTTCGGGTGATATTGGACCGTCTTCTCTATTGTCTGGCTACCATTTGTTGTCTGACCAATGTC  
TTGCATGATGACCTATCTTTTTAAGATGTGGGTGACCACATGTTTGATTACATACAACCTTTCAGCGACGGATGTCTCGGCTCGAAC  
AACGATGAAGGGCGCAGCGAAGTGTGATAAGCATTGTGAATTGCAGAATTCGGTGAACCAATAGGGATTTGAACGTATACACTGCGCTT  
TCGGGATATCCCTGAAAGCATGCCTGCTTCAGTGTCAATATTCTCTCATGCCACTGTCATCTTGTCTGTGCTTTGCACATGCAACT  
GACAGCAGTGTGTGCTGTGTCATTAAAGTGCTCTTCTGCCCCGTGATGCATTGAATCCATGGGTTTTGTGCTCTGTCTGCACTTGA  
TAAACACCTTGTGCTTTGTGTGTAACCTGTTGCTTTGTCTGTGTCATTGATAACCTGTCCCTCTGTCTGCCACTTCAATTTGTGGCTT  
CCTTCCATGACATGAAGTTAGGTAAAGCAAACCCGCTGAATTTAAGCATATCAATAAGCGGAGGTAATCGAATTTCCCGCGGCCGCCAT  
GGCGGCCGGGAGCATGCGACGTCGGGCCCAATTGCGCCTATAGTGAGTCGTATTACAATTCAGTGGCCGTCGTTTTACAACGTCGTG  
ACTGGGAAAACCCCTGGCGTTACCCAACCTTAATCGCCTTGCAGCACATCCCCCTTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCA  
CCGATCGCCCTTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGACGCGCCCTGTAGCGCGCATTAAGCGCGGCGGGGTGTGGTGTGTA  
CGCGCAGCGTGACCGCTACACTGACGCGCCTACGCCCCCTCCTTCGCTTTCTCTGATGCTTCTCGCCAAGTCGCGGCTTCCGTCAGC  
TCTAATCGGGGGCTCCGTTAGGTGCAATTATGTCTTTACGACCTGACCGAACTGGATAGGTAGGTACCTATGTGG

**Table S2.** Statistics of proteins identified in whole lysate (WL) and insoluble cellular fraction (ICF).

No. of Protein Identified	Dinoflagellate Species				
	CCMP113	CCMP1888	ATCI03	CCMP1937	K1
<b>ICF</b>					
Amphiesmal	356	506	439	523	414
Non-Amphiesmal	524	781	687	726	615
Unknown	352	590	408	571	379
<b>Total</b>	<b>1,232</b>	<b>1,877</b>	<b>1,534</b>	<b>1,820</b>	<b>1,408</b>
<b>WL</b>					
Amphiesmal	393	371	451	422	378
Non-Amphiesmal	743	583	813	678	620
Unknown	613	475	550	557	506
<b>Total</b>	<b>1,749</b>	<b>1,429</b>	<b>1,814</b>	<b>1,657</b>	<b>1,504</b>

**Table S3.** Uniprot annotations and amino acid sequences of transporter homologs identified.

Transport Protein(s) Identified in CCMP113	
<b>Best-matched Uniprot annotation of homologs identified (Accession numbers):</b>	
<b>Clathrin heavy chain 1 (P499510, Q2RBN7)</b>	
<p>&gt;Cluster-16584.55282</p> <p>MMQAPMMAAGAQAAPAEPPISLSTTLNLADHGINAQAFRFGNLTMESDRWLSVKDAAADG  SGQVVVIDMHNGNAVNRRPMKAEASLMNPNDNIMALKGSTEGQPGHFVQVFNLDTKEKLG  VYQAPENLVFWRWLSRMLALVGEKNVYHWNLEVANSTPETVFRSGKLAEAGSQVISYA  ANSQLSWCLLTAISTQDQGRITDGNMQLYSVEKKQQQMLEGHAGCFGNITVADGEGPAGL  LAFMERKAGSLQTKLHVMDVTKPRGEGLPFPFKMQSEVAMPPEAPNDFAVLHLSEKHGV  AFMVTKAGYLFVFDIATATMLVRTRVSQDTIFISTYSSLSGGCIFVNRKGAVLTAKVNEP  TIVSYIMNSLVQLSNRQDVAFNLRKFRGLPGADELFRQRFSHYFASGDYKNAALVAAQCK  SGALRSPQTIQQFKSVQAPAGQSSPILHYFSTLLLEYGRLNALESVELARPVVQQQRREL  EKWLKEDKLECSEELGDIVRPLETKFALSILYLRANAHQKAIAAFV<b>ELGQYDQVAAYARKV</b>  <b>GYQADYSSLLQQLSTSP</b>EGATNFAKSLLTGQSGGPPLIDINQVVKVFMQDQNRLETTSI  LLEALKENRPDQAQLQTQLLAMNLQQAQKVAEAIMQNMFTHYDRHYIGQLCEKAGLMQR  ALEHYQDSADLKRVMHLAHQMSPEFLTQYFSRMPDPTALECLYDLLRHNRLNVAQVA  IKYHEQIGAAGKIVEMFESFGSNQGVFFLGAAILSSSTDPLVHYKYIQAASRCGNMQEVER  VCRESTCYDPATVKDFLKDTPRPLIYVCDLHGFVGELETEYLYKNSLMKYIEVYVVK  VNPTNCPTVIGTLIDLDCSEDFIKTLLQNVRAACPIEPLVAEVEKRNRLRVLLPWLEARV  AEGNQDPYLLHNGIAKIYIDTNRDPETFLKNNAFYDSATVGKYCEDKDPHLAFTAYK<b>RAWG</b>  <b>SCDQQLVDVTNRN</b>LLWRLQARYLVERQSPQLWATVLDPENQHRRSVIDMVVSTALPEAKD  ADEVSATVRAFIADLPNELIELLEKIVLHNSNFSKNKNLQNLILTAIKADKSRVMDYI  NRLDNYDGPDIKIALGDPYHLYEEAFLIYKCKGQNAEAMETLLENIESLERAQEFAR  NEGPVWYRLGKAQLENGAVPEAIESYLRAEDASDYQQVIQVAEREENYEELVRFLVMARS  TAKDQLVDTELVAAYAKTDRLAEMEEFVTGTNTANVQQVGDRLFEERAHKAAKLLYQSI  PNNAKLASCHVQLGEFTQAVDAARKANNPRTWK<b>EVNMACVQAQQFRCA</b>EIAGMHIIVHPDH  LEELITQYEAGGHFEELIALLDLSGLTSDRAHVGMYTELAVLYAKYRSEKLMDFIKLNTAR  LNIPKLIHACERHHLWQHVIPLYTHYDEFDSAANTMMAHSPATAFHDQFQIMQKVSNT  LYYRAVQFYLEEQPMQINSLLTTVTVPKVDHARVVQVVRKSGHLSLIMPYLKAVQQHNLQP  VNEALNELYVEAEQHEELRGSIEDFDNIDQIGLAQKLERHELVEMRRIAALVYKKNKRYK  QSIDLSKLDRMYKDAMETARDSGNNDLAESLLRYFVDEDMRECFAACLFCTCYELIRPDVG  VELAWRKGMDFAMPFLIQVLREYTGRIADLDKKTQKKEEAEEKQKSAPNDYVPDYIGPP  MMGPGGLAGFGGLALTSGPAMPVQVQPGFPQPSMMQAGPGMMMPGAGGMR</p>	(Mascot score: 502)
<p>&gt;Cluster-16584.53449</p> <p>MGGPQPPTPPARWLRAQPPPPPRDTRPKPRAMMQAPMMAAGAQAAPAEPPISLSTTLNLA  DHGINAQAFRFGNLTMESDRWLSVKDAAADGSGQVVVIDMHNGNAVNRRPMKAEASLMNP  NDNIMALKGSTEGQPGHFVQVFNLDTKEKLGVYQAPENLVFWRWLSRMLALVGEKNVYH  WNLEVANSTPETVFRSGKLAEAGSQVISYAANSQSWCLLTAISTQDQGRITDGNMQLY  SVEKKQQQMLEGHAGCFGNITVADGEGPAGLLAFMERKAGSLQTKLHVMDVTKPRGEGLP  PPFKMQSEVAMPPEAPNDFAVLHLSEKHGVAFMVTKAGYLFVFDIATATMLVRTRVSQD</p>	(Mascot score: 232)



TIFISTYSSLSGGCIVFNKRGAVLTAKVNEPTIVSYIMNSLVQLSNRQDVAFNLAQRFG  
 PGADQLFQRQFSHYFASGDYKNAALVAAQCKSGALRSPQTIQQFKSVQAPAGQSSPILHY  
 FSTLLEYGRLNALESVELARPVVQQORRELVEKWLKEDKLECSEELGDIVRPLETKFALS  
 IYLRANAHQKAIAAFV~~ELGQYDQVAAYARK~~VGYQADYSSLLQQMLSTSPEGATNFAKSLL  
 TGQSGGPPPLIDINQVVKVFMQNRLOETTSILLEALKENRPDQAQLQTQLLAMNLQQAPK  
 VAEAIMQMMNFTHYDHQYIGQLCEKAGLMSWALEHYQDKADVKRAMLQAHLMTPEALIRF  
 FGRLPPETSLECMYDLMRHNRQNLNVVVQVAIKYNEQIGVMKIVEMFESFASNEGIFYLL  
 GGILSTCTDPDVHFYKIQAAASRCGNMQEVERVCRESSVYDPLKVKEFLKEAKLPDRPLI  
 YVCDLHNFVDELTEYLYKNSLMKYIEVYVVKVNPVKCPTVVGALVDMDCSEDFIKSLLQN  
 VRNACPIEPLVAEMEKRNRRLRVLLPWLEQRVAEGNQDPALHNALAKICIDTNKDPENFLK  
 TNAFYDSATVKGKYCEERDPLHAYTAYKRAWGSCDNQLVDVTNRNCLWRLQAQYLVRQSP  
 DLWALVLNEDNKYRRSVIDQVVSALPECTSADEVSSSTVKAFISADLPNELIELLEKIVF  
 HKSADFSKNKNLQNLILITAVKADTTTRVMDYINRLDNYDGPEIARIALGEPYCLTEEFV  
 IYKKCEMNSEAMDVLNNIGSMERAQEFARCNDSVWYKLGRAQLGGGQVAETIEAHLK  
 AENANDFAQVIQAAEAQGLYDQLVQYLLMARTKAKDQLIDTELVSYSYAKSDRLADMEEFV  
 GGTNTANIQSVGDRLYEESYKAAKILFQSISSNAKLASCHVQLGEFTQAVDAAKKASNP  
 KTWREVNIAACVQAQQFRCAEIAAMYIIVHPDHLEELITQYEAGGHFEELIALLDGLTSD  
 RAHVGMYTELAVLYAKYRSEKLMDFIKLNTARLNI PKLIHACERHHLWQHVIPLYTHYDE  
 FDSAANTMMAHSPTAFADHQQMIMQKVSNTELYRAVQFYLEEQPMQINSLLTTVTPKV  
 DHARVVQVVRKSGHLSLIMPYLKAVQQHNLQPVNEALNELYVEAEQHEELRGSIEDFDNI  
 DQIGLAQKLERHELVEMRRIAALVYKKNKRYKQSIDLSKLDRMYKDAMETARDSGNNDLA  
 ESLLRYFVDEDMRECFAACLFTCYELIRPDVGVELAWRKGMLDFAMPFLIQVLREYTGRI  
 DALDKKTQKKEEAEEKQKSAPNDYVPDYIGPPMMGPGGLAGFGGLALTSGPAMPQVQQPG  
 FPQPSMMQAGPGMMMMPPGAGGMR

**Best-matched Uniprot annotation of homologs identified (Accession numbers):**  
**NAD(P) transhydrogenase, mitochondrial (Q61941)**

(Mascot score: 392)

>Cluster-16584.47837

MRAVVSAGLRRGARQRRFLSDVRVGVLESAGEQRVALVPSVAQKLIKDGAYAVEV~~ESGAGA~~  
~~VAGFADQQYKDAGCAIATRN~~DVIKNSMLFSVNPPPIGELDSMKGKTAVSWVGRRLPDAKE  
~~LLTKAASSGVQLVDLTAVPRITIAQKLDVLSSQAKCAGHRA~~VMEAAHVYGRFFAPEVTAAG  
 KYPPTNTMILGVGVAGLAAIGTSKALGSVVRWDVVDVSDQVQSMGASWITVDFEESGAGE  
 GGYAKESSAEFQQAQKDTFHKLKECEIVITTAIPGRPSVLI E EYMVKDMKPGSVIVDL  
 AAVGGGNCKLTKRDEAFTTE~~ENGVTIIGYTDLAGRM~~AAQSSAMFAQNLANLLGHVSPKEKAK  
 GFFPSIDAAMTAGEDGDIVRSIICTRDGAEVAMPPPPQPTPPKPKAAAVAAAAPKEVSIF  
 RQDMTNALLVTGGNAIMLGMTAVPSHLLGTFTLAGAAGYQVWGVKHALHTPLMSVTNAI  
 SGATAFGGLMLLEHTSGLAYCLAGLATTASAVNIFGGFIVTQRMMLDLFKRPGDEDEDYSALYA  
 IPGAALVAAPLAGLATVESTGVVSQLLCIGAIGGLSSMTTAQMGCKLGMIGMAGGLTTTML  
 GMPASTYPAAAGMLASGAAMGTALGMKVEPIALPQTVAAFHSLVGLAAMITSVASHQHPV  
 PGANLHNVACIVGDYIGGVTLAGSLVAFGKLNGNLGSAPLSLPGKNLLNAAGFAAQVGLGA  
 TFLTAGGELPLWATVALSNAMGYHLVASVGGADMPVCITVLNSYSGWALVAEGFLLDSPV  
 TVIGSLIGFSGAILTKIMCDAMNRDIMNVI FGGIN VAPKKEGETVQK~~EHTETNADGAAAAAL~~  
~~AESQSVMI VPGYMAVARC~~QNSVAAIAQTLVNKGKVRFGIHPVAGRMPGQMNVLAEAGV  
 PYDWVQEMEEVNPDMDSVDTCLIVGANDTTNSGAQEGDHPHLAGMPVIEVWKAKKVI FMK  
 RTMGAGYADVENPVFYKENTDMLLG NATDTTEQIAAKVKELVG

**Best-matched Uniprot annotation of homologs identified (Accession numbers):**  
**ATP synthase subunit alpha, mitochondrial (P80021)**

(Mascot score: 2,254)

>Cluster-16584.49485

MCSPPSPPRQTMASRALLRRLVAGGARGALAPAAQTLAAAPAAASRPLHTSAAALAAKVSP  
 AELTKILSERIANFQDQDVQEVGRVLSVGDGIARIYGLKGVKAGEMVEFSSGLKGMALN  
 LETDNVGVVVFGDDRAIVEGDSVKCTGTIVDVPIDGGLGRVVDALGTPIDGAGPIQTK~~E~~  
~~RRRVELKAPGIIIPRQSVHEPMTTGLKAVDSLIPVGRG~~QRELIIGDRQTGKTAIAIDTILN  
 QRAVNASADTNSHLYCIYVAVGQKRSTVAQLFEILRKNDCLKYTTIVAATASDAAPLQFL  
 APYTGCAAEYF~~RDNGKHAVIFYDDL~~SKQAVAYRQMSLLLRPPPG~~REAYPGDV~~FYLHSRL  
 LERAAMNEETHGGSLTALPIETQAGDVSAYIPTNVISITDGQIFLETIFYKGI RPA  
 VNVGLSVSRVGSAAQIKAMKQVAGTLKLELAQYR~~EVA~~FAQFGSDLDASTQHQLLRGGIL  
~~TEL~~LKQKQFVPMKSQ~~EII~~VSLSWGTGRGYLDKVATKEILRFEALWLEHIKSTRAALLEDIM  
 AKKQLTTEIESGLKSAMEEFLGANEFGRSS

Peptide ion identified in Mass Spectrometry

Amino acid residue labeled with isotopic tag



**Table S4.** Uniprot annotations and amino acid sequences of species-specific surface orthologs identified.

Ortholog(s) Specific to CCMP113	
<b>Orthogroup number: OG0045624</b> <b>Best-matched Uniprot annotation of orthologs (Accession numbers):</b> <b>Isocitrate dehydrogenase [NADP] (P16100)</b>	
(Mascot score: 112; Peptide score: 60.3)	
>Cluster-16584.46230 MDIGGYRLDKAKADAAMNPSATLNTIIANLEGAKKATGSSGAKKATKITYTYTDEAPML ATHAFYPIIQSFCHHAGVPVELKDISVAGRVIAHFPECLTEEQRTGDILSELGLLAQQGQ ANIIKLPNVSASIPQLKECIKELQSQGYKLPDYPEEPKDDKEKDIQARYAKVCGSAVNPV LREGNSDRRAAVPVKEYAFRYPHSMGKWEPSKTHVSCMTDGDYFSHEKSVCIEKACDAR IELVATDGSVKVLKEKVLPQAGEVLDASFMNCKALCSFFEEQIEDAKARGVLFSLHLKAT MMKISDPIMFGHCVKVYFKDVFAYKYKTFAKLGVNANGLGDVYKKIASLPAAEKAAIEA DIMATYEKRGQMAMVDSH <b>RGITNLHVPSDIIIDNSMPTAIRG</b> GGKMWNVDDKEEDFKATI PDRGYGLVFKEVVDFCRKNGAFDPKTMGACPNVGLMAQKAEYGSHPPTFEAPCGGIIRI VANGTNKVLGMGHVSAGDIWRACFTKDAPIKDWCKLAVARCRANFPNNDKPCKAIFWLD PARAHDVVLMQKVRSYLPEFDTAGLDIEIMSPVEAMRVTCQRAKQGLNTITVTGNVLRDY LTDLFPIMELGTSKMLSIIVPMLAGGGMYETGAGGSAPKHVQQMQKEGHLRWDSLGEYLA LACSIADLAEDTKNGKAKALAAALDKAVGTFLVANKNPSRKVKEIDNRGSHYWLARFWAE ELAKQEEEEALREVFTAAAKEFAEKEESVLQDLIACQKGPMIDIGGYRLDKAKADAAMNP SETLNAIIARLEGAKKSSAKKATKITYTYTDEAPMLATHAFYPIIKSFCHHAGVPVELKD ISVAGRVIANFSEYLTTEQRMNDILSELGQ	
Ortholog(s) specific to CCMP1937	
<b>Orthogroup number: OG0081376</b> <b>Best-matched Uniprot annotation of orthologs (Accession numbers):</b> <b>Pyruvate dehydrogenase kinase (P91622)</b>	
(Mascot score: 849)	
>Cluster-5567.50671 MAMHCSSRLQLASCASRVFHGTRRLASAVAPRGRARPDLCRPHDKRNFNWSSRRLFDEPA AQPALGATIASSRSCSSASTGTPEGEFTPGAYRIVNRVAVHRGIGLTENKFAQLHVGES IDVTEVIHLEEEQIRIRARLKEPDGWSISLVDTSDFQWAIIRDLPFDVGAWAACPEVYGA HSRPARQAFEPSQTVWHGDAVAPHGWGVEMVSHAGLKTPSGLNPNQDAFSTTLTSNDWII CVVCDGHGLHGEVVAER <b>RVASALPLILVQLLLAMGPRA</b> ALSAAFAETQAHLESLAREQFH SGATVATCLVHKSRPECWFAHCSDS <b>RTVLGDFASGQAVVCTQDHNPGDAAEFQRL</b> QLAGA KVLTKKEYEGGEIVGRVFPVPGTDAPGLAVSRALGDGCLKTYGVTAEPDVQDVSKAWASC <b>RA</b> <b>PGVLIASDGLWNVIDPEDAVASLARR</b> RRMHLSLRDGAKALCQRAQETMTNLKLDQAFSYC DDVTVALVAPPAILHVDHDDFVQSEIHRFAAKKPHRCSLQEILGMPDVEHLARFIHDEVP VRYAERIRAIELLPGEWQNADLKEVYQRHVQTFRDIRLIQKAESKRSSLSDPLTLVEFDG MIRNAAIKQDQVQLLVARAMYTLRLERPEVFTTEFVDRWLDDFLNYIGTETLLMQYLAR LEGRPNGIIDPRCDVAACVDRVASTVQDVCMDLHQRROPVIVESFSAVEEDQRAPSFYSYI PMFLKYVMMEILKNSCRATLVTCPEYKLPKRPTSVAVCADEHQVAIRVSDRGRGIPFHV NNIWSYMYSTRHKAGVGFLGATPLAGYGIGLPLSRLYVRYLGGSLGVVSWPGLGCDVHL SLPRLSSEQVEVVPDQDN	
Ortholog(s) specific to K1	
<b>Orthogroup number: OG0038250</b> <b>Best-matched Uniprot annotation of orthologs (Accession numbers):</b> <b>Fucoxanthin-chlorophyll a-c binding protein, chloroplastic (Q39709)</b>	
(Mascot score: 370)	
>Cluster-24592.102021 MAQGCAFSCVDCHNQFLRLSFLSSLTMQKVAIILGCLVCTGYGKETKKALKKPSDSLAKLL LANRREAAFNASAGQAPLASSRPRSQGSMMATS <b>RSLPWME</b> SPAHLEGMIGNAGFDPLGIST <b>PENVVRW</b> MREAEKHKGRMCMLAWLGVAVDVGIKFPGERY <b>EVLTSFNAHDATVEKEL</b> FLLL LNWGTF <b>ETIGFNQILDMMLDPTNPRQ</b> AGDYGFDPMGIIITPENEEWYKTAE LTHARLAML F	
<b>Orthogroup number: OG0038286</b> <b>Best-matched Uniprot annotation of orthologs (Accession numbers):</b> <b>Fucoxanthin-chlorophyll a-c binding protein B, chloroplastic (Q40296)</b>	
(Mascot score: 777)	

>Cluster-24592.122404

MDSRAYLAPSWLFFVSSHTVHPLPPTMQKAAFLASLVCVGAFTPTPQQFRPVQQRSAGQ  
PSASRQSVAMADSASGLSWWKPVGNTWDPLGLAKSPEKFERLRYVEVKHG**RISMLAVLG**  
**HIVAASGARW**PGQLYGGLKFADVKSGFKALSQSPQDYAVIVAFVAFLELRVMKEQVKG  
EFPDGL**RNGLFKDGLDTEKEK**ASKINKELNNG**RAAMMGIFGLMVHEM**LNGKPYVINEL  
LGMGPY

(Mascot score: 409)

>Cluster-24592.47428

MTASNTFTMKNFALVLACLVLGQSFTPTVQQLRPSQHGPAAGRSSSKQLITMANSAPGL  
SWYKAGIGNTWDPLGLAKTPEKFERLRYVEVKHG**RIAMLAVLGHIIVTATGYRW**PGELYGH  
LKFSQVKGSGFKALSQSPQDYAVILAFVAFLELRVMKEQVKGEFPGDLRNGLFKEGWDV  
LSEKEQVSKINKELNNG**RAAMMGIFGLMVHEM**LNGKPYVINELLGWGPY

(Mascot score: 325)

>Cluster-24592.64372

MVASPHTVPSFLPTMQKVAIALASLACAGQAFTPTPQQFRPSQQRSAQAQPSASRQSVAMA  
DSVPGLSWWKPVGANTWDPLNLAATPEKFERLRYVEVKHG**RISMLAVLGHIIVAASGYRW**P  
GGQFYGGGLKMAVDKSGFKALSQSPQDYAVILAFVGFLELRVMKEQVKGEFPGDLRNGL  
FKEGWDLLTEKEKASKINKELNNG**RAAMMGIFGLMVHEM**LNGKPYVINELLGWGPY

(Mascot score: 306)

>Cluster-24592.96057

MSQLDPKDYAVILAFVAFLELRVMKEQVKGEFPGDL**RNGLFKDGLDVLSEKEK**ASKINKE  
LNNG**RAAMMGILGLMVHEM**LNGKPYVINELLGWGPY

Orthogroup number: OG0019834

Best-matched Uniprot annotation of orthologs (Accession numbers):

Fucoanthin-chlorophyll a-c binding protein F, chloroplastic (Q40300)

(Mascot score: 877)

>Cluster-24592.21292

MVAACVSGLLNYSTSI CNLIMRAQVLVFFCLACAGQARRVQMPKEHTQGA AVKELLRALK  
PEAGFQAQMTQPTRIGELAARKASAPTMYDLTGEIGATGPLL PYWDPLGLAKDATPEQFR  
RWRVELKHGRIAMVATLGYAVQEVLRWPGYLSPSAGVKFADIPNGIRGLAAVPPLGLAQ  
ILLAI GLMEVATWRYEYEGPWPGSV**EGKEPGDVAGDLWVRY**TDPEEKYKLNIELNNG**RA**  
**AMMGS LGMLMHDHITG SWIPPGFK**

Orthogroup number: OG0014087

Best-matched Uniprot annotation of orthologs (Accession numbers):

Dual specificity mitogen-activated protein kinase 4 (A0A5C6NSI7)

(Mascot score: 454; Peptide score: 150.99)

>Cluster-24592.69574

MGIEVLGVVVASLSDPGHSMQEDSYSII IQRNAEMEHMNVLDLEDVGNKDDGGCARREET  
MPLMDHVHVEEVIDREM QDNKGDAIDVRLYRLRMGAIMACAFGMVALATGIFAYHNSMSS  
SRRPQDSMRFFLSNGQCKRPAGWCMSNPCFTGNYSLYDCDGDGQLDPFCWPALAQPGDRL  
CGCLQSSNGCIDTKLFNFLCPAPTTAAPTAPTPAPTAPTPAPTAPTTAAEPTMTLP  
TPYIPVVPAPKESILA AWGDDIKVDYCVPKGHLVGTFNITGKMIDVKMPEDGCITAKAPKV  
VNDPIYEGYQLIKYRHPKSVDLSSSDMKVKSPPYPGIMKEWLKSVGAKVDPGDAVARIFV  
NGDVKTIA YTDGPSGVVDRAPFLPSDPVAAGTDLIVVALGIFVKWEADVGEHVKKDDLI  
CSMTQGYIVKPDGIPADTAANDEILGWVAANSKKLFPKSTIPLLASKNGSVVQRSDEEPG  
DIIGINRKTPTGALILSHLPPVPVPASYISRYNSTPPNVTKAPDGAGVTPKWYSSDGSKY  
LYTGAWLKLKAWKVESGDQVKSGEMVATAENLEGVSSNLTRYSGYIVNKQYINPGDFIG  
PSDELVLVGNYLAPAAGSSFPWWVLLPLCLLCLIPLCCKSEPPPPPPPPPPPPPPVVV  
EPAPAPVVVAPPPPKPEPKPEPKPEPKPEPKPEPVVPPPPPPPGFPIYFDGKEYRCEYHP  
LGIKFWQKAPIKVEEYNFNSYKTLGITKGMTLTKIKDTEIRPDQKFKDVHLLLEAMAP  
LPVWPLHMGFKTPGGELKEYWFNEHPIGIEFTRQAPIKIDTVKSGSLAAEVGAQPGWTIV  
KLGDSDISGDHNF SHVSKLLKDGHLNLP**RGK****AQMGTYADNLGAGSHSGGGVTEAPRPST**  
EAPAGFK

(Mascot score: 804; Peptide score: 128.21)

>Cluster-24592.107035

MTMRIAILVFACLALTDAQRVRLTGEPREATTQALSALLQTRKPTAGFQPSGNVLGNTPR  
SPSSKLAERPAQPMQVDTLGQLGATGPLL PYWDPLGLAKDATPEQFRWRVELKHGRIA  
MAATLG YFVQEVLRWPGYLSPSAGIKFSDIPNGIRGLAAVPPLGLAQILLGIGLMEVATW  
RYYEGPWPGSVPEGKEPGDVAGELWVRYSDPEEKTKLNIELNNG**RAAMMGS LGMLMHDH**  
**ITG SWIPPGFK**

---

**Orthogroup number: OG0019848**

**Best-matched Uniprot annotation of orthologs (Accession numbers):**

**CAAX amino terminal protease family protein (M2X807)**

**(Mascot score: 268)**

>Cluster-24592.117554

MSWLLKIDRVTFISFFTMRLALLLVCLACVGHGRRVYTSPGDSHSSALAGRQKTSKGIDV  
 DSRRLGAPIASRSLAALLRAGSPAAGITASSRPTLPHLPSRTAATPPRSLRSVLGLRGGR  
 STGGEIEWPEEELEDDGMAIDVERGSSDVASEQEVVLDENDATKAEESSDSSPSEDSDAV  
 YYKGGASGELTEENVLPTLVQTPILIAAYLLHLTWVSKNQLPLGR**RSLPYENIVGAVAVS**  
 NYFVSQIRKWWLRRQQGLEATPTTVPWGKVHNGPFSLLLTLMCMGAHWLSNYIGIAFDY  
 FLLFLENRGVPMTWGQAYAMQQLFSLHAYVAMAVHAMGSRLKPFPPPPFGKGEWIRVRWH  
 TNWLGWVVGGFYASTLACNSLTASCHALYNYLVKNAGLPPAPTQVPLTETVNVNQLSRPG  
 GGDIAAFLVGLLAPCVSAPIFEEVLYRGFLLPTLH**RFMPLGAALPLQGLLFGLHHRA**IPT  
 LLPLSLLGWLWGWFYVKTGNLVIPMLMHAMWNSRSFVNSLLEKAY

---

**Peptide ion identified by Mass Spectrometry**

**Amino acid residue labeled with the isotopic tag**

**Table S5.** Uniprot annotations and amino acid sequences of STX-producing-species-specific surface orthologs identified.

Orthogroup number: OG0034291	
<b>Best-matched Uniprot annotation of orthologs (Accession numbers):</b>	
<b>F-type H<sup>+</sup>-transporting ATPase subunit beta (A6BM09, Q06J29)</b>	
<b>Ortholog(s) of CCMP113</b>	(Mascot score: 2,553)
>Cluster-16584.51079 MKCKVDFEFKFMMTQIRLVQRYVFLPSVVDALLVMRPSCNFRDGTVVKNVCVANFGQYFVE LPFNDSYLFGLIHLIPGDYKSTLHCFYSSMCREVCNVRDIVLGFASVFLFVKSYANCLVA ELSQLCYGGVLRVALGSTDGLCTWRCTFLMVLQPVIVPVGRIALGRIFNVVGSVIDRYM ELSLSSQFNVSVPIDNGLFVESCEDLTYYTLGYPDAITLVHTKALDLNLISLWNTMNVST ISKSWIYYIGCLYQLVISQVCRHLHQNLYGTHESVIFEEIDRSFNALS RACESLFYNTDT LFALVKPIHKTPVAIMTSLIHLTLFETGIKVVDDLTPYKKGKIGLFGGAGVGKTVVIME LIRNLAVEHGGLSLFAGVGERTREGNDLYCEMQDSGIIISLTLRQGKSQSSVCYPCLMYQP LFAAAWSQVVLVFGQMNETPGSRMRVTHASLAAAEYFREAFQDVLIFVDNVFRFLQAGS EVSTLLGRMPSAVGYPPTLSTEMASFQERIVATTAGSITSIQAIYVPADDLTDPAVPVIF GHLDVAVTVLSRALASKGIYPAVDALNSTSKMLDPSYVKQEHFCVATDVKQMLQRYKELQD VIAILGLEELSDQDRIVVDRARKVERFLSQPFFVAEVFTRIQQGRYVSLNDTIIGFSQIVQ GELDEVSEGGFYLGKALSDVVHDNDA	
<b>Ortholog(s) of CCMP1888</b>	(Mascot score: 1,414)
>Cluster-6994.52323 MKCKVDFEFKFMMTQIRLVQRYVFLPSVVDALLVMRPSCNFRDGTVVKNVCVANFGQYFVE LPFNDSYLFGLIHLIPGDYKSTLHCFYSSMCREVCNVRDIVLGFASVFLFVKSYANCLVA ELSQLCYGGVLRVALGSTDGLCTWRCTFLMVLQPVIVPVGRIALGRIFNVVGSVIDRYM ELSLSSQFNVSVPIDNGLFVESCEDLTYYTLGYPDAITLVHTKALDLNLISLWNTMNVST ISKSWIYYIGCLYQLVISQVCRHLHQNLYGTHESVIFEEIDRSFNALS RACESLFYNTDT LFALVKPIHKTPVAIMTSLIHLTLFETGIKVVDDLTPYKKGKIGLFGGAGVGKTVVIME LIRNLAVEHGGLSLFAGVGERTREGNDLYCEMQDSGIIISLTLRQGKSQSSVCYPCLMYQP LFAAAWSQVVLVFGQMNETPGSRMRVTHASLAAAEYFREAFQDVLIFVDNVFRFLQAGS EVSTLLGRMPSAVGYPPTLSTEMASFQERIVATTAGSITSIQAIYVPADDLTDPAVPVIF GHLDVAVTVLSRALASKGIYPAVDALNSTSKMLDPSYVKQEHFCVATDVKQMLQRYKELQD VIAILGLEELSDQDRIVVDRARKVERFLSQPFFVAEVFTRIQQGRYVSLNDTIIGFSQIVQ GELDEVSEGGFYLGKALSDVVHDNDA	
<b>Ortholog(s) of ATCI03</b>	(Mascot score: 2,425)
>Cluster-15238.40798 MRNHGSVCISVGPVIDVMKCLVFNFDFKMITQIARPSGFVVESNVFLPSVYDAILVIRP SCNFRDGTVVNNCIVNYGQYLCELPFNDSLLGLIHLPLVQVKYLTISFWSNGSSDTHS AMCKEACEMNIRDILFGFASVFLFIKSYANCLVAELSQLCYGGVLRVALGPTDGLCTWG CSFLMVLQPVIVPVGRIALGRIFNVVGSVIDRYMELSLSCQFNTAVPIDLGLFVESHPEL TYALSYPDITLTLAHTKALDFSLVSHWNTILSSNTISSAWVYYIAYLYQLVMNHVHLNVK DRTQKIKIS EIDCSVDALSRACESLFYNTDTL FALGKPIHQTPVAIMTSLIHLTLFETGI KVVDDLTPYKKGKIGLFGGAGVGKTVVIMELIRNLAVEHGGLSLFAGVGERTREGNDLY CEMQDSGIIISLTLRYPDPHTPRIMYQPLFAANQSQVVLVFGQMNETPGSRMRVTHASLAA AEYFRDAFGQDVLIFVDNVFRFLQAGSEVSTLLGRMPSAVGYPPTLSTEMGSFQERIVAT NAGSITSIQAIYVPADDLTDPAVPVIFGHLDVAVTVLSRALASKGIYPAVDPLNSTSKMLD PSYVKQEHFCVATDVKQMLQRYKELQDVIAILGLEELADQDRIVVDRAARKVERFLSQPFF VAEIFTRIQQRYVSLLDITIRGFSQIIKGELDQVSEGGFYLGKALSDVTFGSECIRHVISS NTNVPNT	
<b>Ortholog(s) of CCMP1937</b>	(Mascot score: 1,463)
>Cluster-5567.44538 MITQIRPVIADQQLHPLVRDVFFPSVVDALLVITSQCNFRDGTVIYNVGSNQAQYFAELP FNDAYSVALIHLCCAVFDSRSPVARGCTPCGVHEYIGMTQISSEIWDNMNARDVVLGFAAL FLLVKSANCLVAELSQLWQGGVMRAIALGSTDGLSCFSCTFLTLTFQPVIVPVGRIITLGR ILNVVGSSIDPYIHLSSQFGPYHLVSAELFVLESSLESSSYATTFISEGCLALDHSLIS LYTTMNVSTISFPWIYYLACLGRAMVRCMAVCVPPEVGLLDMLESCRSTDGHYKTIDRS VEVIQHCNINIFYHADTLFASIKPIHGTPLAINELRICVILFETGIKVVDDLTPYKCGGK IGLFGGAGVGKTVVIMELIRNLAVEHGGLSLFAGVGERTREGNDLYCEMQDSGIIKISLS	

---

IPKEQYSLNPEGSSYSVVTRL SQVVCVFGQMNETPGSRM RVTHASISMAEYFRDAFSQDV  
LIFVDNVFRFLQAGSEVSTLLGRMP SAVGYQPTLSTEMGSFQERICARLTGSITSIQAIY  
VPADDLTDPAPVLI FGHL DATTVLSRALASKGIYPAVD PFNSTSKMLCPECVKQEHFCVA  
SDVKQMLQRYKELQDLIAILGLEELSDQDRIVVDRARKI ERFLSQPFFVAEVFTRI SGRY  
VSLNDTIRGFTKIVTGELDAQSEGGFYLGAITDVM

---

Peptide ion identified in Mass Spectrometry

Amino acid labeled with isotopic tag