

***New Phytologist* Supporting Information Figs S1–S4 and Table S1**

Article title: The phycocyanobilin chromophore of streptophyte algal phytochromes is synthesized by HY2

Authors: Nathan C. Rockwell, Shelley S. Martin, Fay-Wei Li, Sarah Mathews, and J. Clark Lagarias

Article acceptance date: 4 December 2016

The following Supporting Information is available for this article:

Fig. S1 Gene models for Notes S1.

Fig. S2 Synthetic gene sequence for the catalytic core of KflaHY2.

Fig. S3 Synthetic gene sequence for the catalytic core of KflaHY1.

Fig. S4 Synthetic gene sequence for the photosensory core module of MvirPHY1.

Table S1 Accession information for FDBR and RCCR sequences

Notes S1 Flat text alignment of FDBR sequences.

Notes S2 Formatted input file for FDBR phylogeny.

Notes S1 and S2 are supplied in a single compressed archive (MD5 checksum 9bb79dbcd9e86431f1d13e56752a40f0).

(a)

CCMP329_PcyA

underline = EST EC654063 (Genbank)

cyan highlights = genomic Contig25718 (Cyanophora genome home page)

LIKRLADTIEEVWENKLDLTPYELPEDLGIVEGRLEGERLVIENKCFOSTAFRKMHELELARVGNLNDILHC
VMFPNKEYALPMFGADLVGGRGQLSACIVDLSPTSPNGLPAGYVERLEGLERPQFSQPRELPPWGHIFSPF
CTFIRPAGENEEDQFLALVRSYMEIHCDMANIMPKLTAVDDINNNLRGQKSYCTSQQQNDKTRRVLEMAFG
DAWAERYMTTVLFD

(b)

SL-175_PubS

green highlights = large insertions (***) in Supp. Note 1)

VWHAHMSWLFARLEDALGPLTAVNIPDDLAVQTRPNKAHAQTWVFSSPHARRIRFTYVDAGVNAQIFNCVV
YPRC **DAEENDAQGTTAATSPSKGTATNESNGITSTVAAAT**CLGDAPLLGVDLLCLAGGRKILVGIDLQPLS
RGDAYLARYAPALAAATREERFPDLNLVQPSKKFYEDAQYFSPAMLFARPEAAEMARA **AAAAAAAAAGADFP**
VCLDTCADHDCGVELLQARVLDVAVKVYTEEFLRVLDV **EEDASAGRGEARARGGGFGNNVINSIMASVGVG**
VAMKRGTDEAGEVAAAETAAREGADTEVSHVPQLEPQLEPQLEKPEPESPTLVAWDDGEAAVEAVAAGE
MENTTARAVARAAVEEATGARGKGGSRSSPQDAAVVVAAAASDAAEQQPQAEAESPTLVSWDDGEAAVE
AVSAGEMENTTAHAAAARAAVEEAAGVRGKRSPRARILSPMATAAAQDSHDDWQRARDPAVPMFASYFGKEW
ANRMAHEVLFP

Fig. S1 Gene models for Notes S1. (a) For *Cyanophora paradoxa* CCMP 329, neither the draft genome (cyan) nor the available expressed sequence tag (EST) data (underline) contained a complete PCYA gene model. However, the available sequences overlapped each other, permitting inference of a complete sequence for the catalytic core region used in our phylogenetic analysis. (b) For *Mantoniella antarctica* SL-175, the PUBS sequence contained large insertions (green) that were replaced with asterisks in the structurally corrected alignment (Notes S1) and with gap characters in the input file for phylogenetic analysis (Notes S2).

```
ATGGAAGAAGCAAGTGC GGGTGAATCGGGCAGCAATGATGGTCTGTACAAACCGTTTCGTTGAATACGCCGTGAAGGCTCTGG
M E E A S A G E S G S N D G L Y K P F V E Y A V K A L
GTGAAAAAATGGAAGTGTGGCCGTATCCGATTAAGAAGAATTTATCCGCCTGGAAGCGACCGATTCCAAGGGTCAGAAGGCCGT
G E K M E L W P Y P I K E E F I R L E A T D S K G Q K A V
GACCGACAGTACGTACCTGCAGTCCATAAACTGCGTCAAATGCGCGCGTGCAACATTGAGCCGGCAACAATCTGCAAGTTCTG
T D S T Y L Q S H K L R Q M R A C N I Q A G N N L Q V L
AATCTGATCGCGTTTCCGCGCCCGAATATGATCTGCCGTAATTCTGTGCAGACCTGGTACCCTGCCGCGTGGTCATCTGCAAA
N L I A F P R P E Y D L P Y F C A D L V T L P R G H L Q
TTCTGGATCTGAACCCGCTGCACAAAACCCCGGAACATTTTAAAAACACATCAAGCCGCTGCTGCCGATTACCAATCGCTATAT
I L D L N P L H K T P E H F E K H I K P L L P I T N R Y I
CAAAGTCTGCCGTGGGGCGGTGAATTCACGGCAGAATCACTGCAGTTTTTCTCGCCGGCTCTGCTGTGGACCCGTCAGAGCAT
K V L P W G G E F T A E S L Q F F S P A L L W T R P S D
CGTACCACGGTCCAAACGGTATCTTGAAGCCTTCAAGGAATACATCGATGCGTACGCCGAAATGGTTGCAAAGGCTGTCCGA
R T T V Q T V I F E A F K E Y I D A Y A E M V A K A V P
CCACGGACCCGGTCAAATTCACGTAACCCGGAAGCTCAGCATAAATATGTGTGCTGGCGTGCAGAAAAAGATCCGGGTCGTCC
T T D P V E I A R N R E A Q H K Y V C W R A E K D P G R P
GACCATTACGCGCTGTATGGCACC GAACTGTGTGAAGCGTACATCCACGGCTTTCTGTTGACGGTATGAAAGA ACTGGGCAGC
T I T R L Y G T E L C E A Y I H G F L F D G M K E L G S
AAGTCTTTTCTGGATTATTTCCCGAATACCGTGGCCCGGACGGTAACGTGATTA AAAAGCGCTCTGTGGTTGGTAAAAAGTGGC
K S F L D Y F P E Y R G P D G N V I K K R S V V G K K C
TTACCCGTCGTTGGGATCTGGACGGCAATTTTATCCCGCCGACGTAA
V T R P W D L D G N F I P P T
```

Fig. S2 Synthetic gene sequence for the catalytic core of KflaHY2. The synthetic gene (Genscript) and coding sequence are shown.

CATATGGAGGAGAAGAAAGGTCGTCGTCGTCGCGGGCGAAGCGAAAGGTTTTGTGGAAGAGATGCGTATCAAGGCG
GTATACCTCCTCTTTCCAGCAGCAGCAGGCCCGCTTCGCTTTCCAAAACACCTTCTCTACGCATAGTTCCGC
M E E K K G R R R P G E A K G F V E E M R I K A

ATGAAGCTGCACACCCGCGTCAAGCGAAAGCGGCGAGGCGCAAGAGAAAGAAACCCAGCAAAAACCGCTGCCG
TACTTCGACGTGTGGGCGCCAGTCCGCTTTCTGCCGCTCCGCGTTCTTTTCTTTGGGTCGTTTTTGGCGACGGC
M K L H T R G Q A K D G E A Q E K E T Q Q K P L P

CAGTGAAGCCGACCGTGGAGGGCTACATCCAATTCTGGTGGACAGCAAAGCGGTTTATGATACGATGGAGGGC
GTCACCTTCGGCTGGCACCTCCCGATGTAGGTTAAGGACCACCTGTCGTTTCGCCAAATACTATGCTACCTCCGC
Q W K P T V E G Y I Q F L V D S K A V Y D T M E A

ATTGTTGACAAAGCGCCGACCCAGAGCTACGCGGAGTTCGGTAACACCGGTCTGGAACGTACCGGCCCGCTGAGC
TAACAAGTGTTCGCGGCGTGGTCTCGATGCGCCTCAAGGCATTGTGGCCAGACCTTGCATGGCCGGGCGACTCG
I V D K A P H Q S Y A E F R N T G L E R T G P L S

AAAGATCTGGAGCACTTTCTAGCGAAGGTTATAGCATCCCGGAGCCGACCGAAGCGGGTCTGACCTATAGCAAG
TTTCTAGACCTCGTGAAGCATCGCTTCCAATATCGTAGGGCCTCGGCTGGCTTCGCCCAGACTGGATATCGTTC
K D L E H F R S E G Y S I P E P T E A G L T Y S K

CTGCTGGAGGAACTGAGCGAGAGCAACCCGCAAGCGTTTCATTTGCCACTTTTACAACGTGTATTTTGGCGACAGC
GACGACCTCCTTGACTCGCTCTCGTTGGGCGTTCGCAAGTAAACGGTGAAAATGTTGCACATAAAACGCGTGTCG
L L E E L S E S N P Q A F I C H F Y N V Y F A H S

GCGGGTGGCCGTATGATCGGTCTGAAAGTTAGCGAGATGATTCTGGAAGCCGTGAGCTGGAATTTTATAAGTGG
CGCCACCGGCATACTAGCCAGCATTTCAATCGCTCTACTAAGACCTTCGGCACTCGACCTTAAAATATTACC
A G G R M I G R K V S E M I L E G R E L E F Y K W

GACGGCGAGCTGAGCGAACTGCTGGATAAGGTGAAACTGAAGATCAACGAAGTTAGCAACGACTGGACCCGTGAG
CTGCCGCTCGACTCGCTTGACGACCTATCCACTTTGACTTCTAGTTGCTTCAATCGTTGCTGACTGGGCACTC
D G E L S E L L D K V K L K I N E V S N D W T R E

GAAAAGATAAGTGCCTGCAAGAGACCGAACTGAGCTTCAAGTATAGCGGTAGCCTGCTGAAACTGCTGGCGAAC
CTTTTCCTATTACGGACGTTCTCTGGCTTGACTCGAAGTTCATATCGCCATCGGACGACTTTGACGACCGCTTG
E K D K C L Q E T E L S F K Y S G S L L K L L A N

ACCATTTACCTAGCGTGACAACCTTATCCTCGAG
TGGTAAATGGATCGCACTGTTGAATAGGAGCTC
T I Y L A

Fig. S3 Synthetic gene sequence for the catalytic core of KflaHY1. The synthetic gene (Genscript) is shown with the open reading frame in green and with flanking sequences in red.

ATGGGTACGCTGTCAGGCTCGCAGGCAAACAGTGTGGAAGTTGCACGCGCTCAAGCGGATGCCGACGCAACCCTGAATGATGC
 M G T L S G S Q A N S V E V A R A Q A D A D A T L N D A
 GTATGAACAGGAAAAATCTCATGGCGTCCACTTTGATTACGGCCATAAAGTGTGGACGCTAACAGAGCGCGCGTATGTCTG
 Y E Q E K S H G V H F D Y G H K V L D A N Q S A R M S
 AAAGCTCTGGCGGTGGCACGAGCTTCGGTGCTATGGCGTATGCAACCGTATGCAGCGTCCGGGTATTATCCAATCATACGGC
 E S S G G G T S F G A M A Y A T R M Q R P G I I Q S Y G
 TGCTGTCTGGTGGTTGATGCGGTGTTTAAAATTATCGCGGCTCGGAAAACGCACTGGCTATGCTGACCGGTAAAACGGTTCA
 C C L V V D A V F K I I A A S E N A L A M L T G K T V Q
 GGATAAACATGCCATTATCGGCCAAGACAGCCTGTCTCTGTTTTCCGTGGCAGCTGCGGCCACCTGAAGAAAACGGCGTAGTT
 D K H A I I G Q D S L S L F S V A A A A H L K K A A S
 CCAAAGATCTGGGTGTCGTGAATCCGGTTGCTGTGTTTCCGCGGGTACCGGTAACCGTTTTTTCGCGATTATGCATCGCAAC
 S K D L G V V N P V A V F A A G T G K P F F A I M H R N
 GATGTTGGCCTGGTCATTGACCTGGAACCGATTACCCGCGAGGCAAATACCGGTCCGGCCCAAGCGGTGCACTGGCTTCTCA
 D V G L V I D L E P I H P Q A N T G P A Q A G A L A S H
 TAAACTGGCCAGTGATTCCGTGGCAAACTGCAGTGC GTTGGTCCGGCAACCTGGACCAAGTGTATAAAATTCTGGTGAAG
 K L A S D S V A K L Q C V G P G N L D Q V Y K I L V E
 AAGTTCGCTACTGACCGTTATGATCGTGTATGGTCTACAAATTTTCATCCGGACATGCACGGCGAAGTTGTCGCTGAATCG
 E V R S L T G Y D R V M V Y K F H P D M H G E V V A E S
 GTTGATGCGGGTGTGCAACAGTTTTCTGGGCTGCATTATCCGGTACCGACATTCCGCGAGCAGCTCGTGCCTGTTCTGTCG
 V D A G V E Q F L G L H Y P A T D I P Q A A R A L F V R
 TAATCGCGTTCGATGATCGCGGATGTTACCACGAAAGCGTCCCGATGCTGCAGGATTCTTACCTGAAACACCCGGTTGACA
 N R V R M I A D V T T E S V P M L Q D S Y L K H P V D
 TGTCATTCTCGACCCTGCGCGGTGTGCTGCATTGTACCGTGAATATATGGCGAACATGGGCAGCCTGGCCTCTCTGGTTATG
 M S F S T L R G V L H C H R E Y M A N M G S L A S L V M
 GCAGTGGTTGTCAATGAATACGATGAAAGTGCATCGTCCCCTGCCCCGATTGGTACGCGTCTGTGGGGCTGGTGGTTTG
 A V V V N E Y D E S D R T S R R P I G Q R L W G L V V C
 CCATCACATTACCCCGCGCTGTGCTGTATCCGCTGCGTAGTGCCTGTTCTTTCTGGTGCAGTTTTTACGCTGCAAAATTA
 H H I T P R V V L Y P L R S A C S F L V Q V F S L Q I
 ACGCAGAAATGGAACGACGTTTTAGCAACGTGAAAAAGAAGTGATTCGCACCCAGGCCATCCTGTGCGATATGCTGCAACGT
 N A E M E R T F Q Q R E K E V I R T Q A I L C D M L Q R
 GACGACCCGATTGGTATCGTTACCCAGACGCCAATATCATGGATCTGGTGCCTGTGATGGTGCGGGTCTGTGGTATGGTGG
 D A P I G I V T Q T P N I M D L V P C D G A G L W Y G G
 CCGTCTGCTGACTGGGCGTTACCCCGTCCGAAGATGCAGTCAACAGCTGGTGAATGGATGCTGATGCATCACGGTGATA
 R F W S L G V T P S E D A V E Q L V K W M L M H H G D
 GCACCGGCTGTCAACGGACAGCCTGACCGAAGCGGGTTATCCGGGCGCCGAAGCGCTGGCATCTACCGTGTGTGGTCTGGCC
 S T G L S T D S L T E A G Y P G A E A L A S T V C G L A
 GCATGTCGATTACCGAAAACGATTTTTCTGATGTGGTCCGCGGTCCGACCAACCGTACGATCAATTGGGCGGGCCAGACCCA
 A C R I T E N D F L M W F R G P T N R T I N W A G Q T H
 TGATCCGAATGCAGCTGATGACCAGAAGAAAATGCACCCGCGCTAGTTTTGCGGCCTTCAGGAAGTTGTCCTGGTCAAT
 D P N A A D D Q K K M H P R A S F A A F Q E V V R G Q
 CCGCCCCGTGGGAAGATGGCGAATGGACATTATCCATTCTGCGAGCTGATTCTGCGCGGTCAATTTGCAGATATCGAAGAC
 S A P W E D G E M D I I H S L Q L I L R G Q F A D I E D
 GTCGCGTGCCTACGATGATTCACTGCGCCTGAATCAGTGCCTGATAAAACATCGATGAAGTCAATACCGTGGCGAGTCC
 V G V R T M I H L R L N Q L R L K N I D E V N T V A S P
 CGGG
 G

Fig. S4 Synthetic gene sequence for the photosensory core module of MvirPHY1. The synthetic gene (Genscript) and coding sequence are shown.

Table S1 Accession information for FDBR and RCCR sequences¹

Protein	Organism	Database	Accession
PcyAANASP	<i>Nostoc</i> sp. PCC 7120	DOE-IMG	alr3707
PcyANOSPU	<i>Nostoc punctiforme</i> PCC 73102	DOE-IMG	Npun_R5577
PcyATRICH	<i>Trichodesmium</i> <i>erythraeum</i> IMS101	DOE-IMG	Tery_4156
PcyASYNP7	<i>Synechococcus</i> <i>elongatus</i> PCC 7942	DOE-IMG	Synpcc7942_1225
PcyASYNY3	<i>Synechocystis</i> sp. PCC 6803	PDB	2D1E
PcyAFREDI	<i>Tolypothrix</i> sp. PCC 7601	DOE-IMG	fdiDRAFT29270
HLUCCO16_03155	<i>Phormidium</i> sp. OSCR	DOE-IMG	HLUCCO16_03155
Lepto7104D_5008	<i>Nodosilinea</i> <i>nodulosa</i> PCC 7104	DOE-IMG	Lepto7104DRAFT_5008
PcyACROWA	<i>Crocospaera</i> <i>watsonii</i> WH 8501	DOE-IMG	CwatDRAFT_2748
PcyASYNP2	<i>Synechococcus</i> sp. PCC 7002	DOE-IMG	SYNPCC7002_A2228
PcyAGLOVI	<i>Gloeobacter</i> <i>violaceus</i> PCC 7421	DOE-IMG	gvip359
PcyASYNEL	<i>Thermosynechococcus</i> <i>elongatus</i> BP-1	DOE-IMG	tl12308
PcyASYNPX	<i>Prochlorococcus</i> sp. WH8102	DOE-IMG	SYNW1084
PcyASYN9605	<i>Prochlorococcus</i> sp. CC9605	DOE-IMG	Syncc9605_1216

PcyASYN9902	<i>Prochlorococcus</i> sp. CC9902	DOE-IMG	Syncc9902_1253
PcyAPROMA	<i>Prochlorococcus</i> <i>marinus</i> CCMP 1375	DOE-IMG	Pro0819
PcyAPROMM	<i>Prochlorococcus</i> <i>marinus</i> MIT 9313	DOE-IMG	PMT0590
SynRCC307_1284	<i>Synechococcus</i> sp. RCC307	DOE-IMG	SynRCC307_1284
CYME_CMG110C	<i>Cyanidioschyzon</i> <i>merolae</i> 10D	Genbank (genomic)	CYME_CMG110C
SAG4_97_PCYA	<i>Cyanoptyche</i> <i>gloeocystis</i>	MMETSP1086	36603_1
SAG46_84_PCYA	<i>Gloeochaete</i> <i>wittrockiana</i>	MMETSP transcriptome	177546_1
CCMP329_PCYA	<i>Cyanophora</i> <i>paradoxa</i>	multiple sources	See Fig. S1
CCMP717_PCYA	<i>Nephroselmis</i> <i>pyriformis</i>	MMETSP0034	35417_1
CCMP1998_PCYA	<i>Pycnococcus</i> sp.	MMETSP1085	7414_1
SL-175_PCYA	<i>Mantoniella</i> <i>antarctica</i>	MMETSP1106	20204_1
RCC2596_PCYA	<i>Ostreococcus</i> <i>mediterraneus</i>	MMETSP0932	4952_1
CCMP2329_PCYA	<i>Picochlorum</i> <i>oklahomensis</i>	MMETSP1161	3223_1
CCMP1320_PCYA	<i>Dunaliella</i> <i>tertiolecta</i>	MMETSP transcriptome	38532_1
CCMP219_PCYA	<i>Chlamydomonas</i> <i>euryale</i>	MMETSP0063	48966_1

PCYACHLRE	<i>Chlamydomonas reinhardtii</i>	Genbank (genomic)	XP_001693622
PebAGLOVI	<i>Gloeobacter violaceus</i> PCC 7421	DOE-IMG	gvip179
PebAFREDI	<i>Tolypothrix</i> sp. PCC 7601	DOE-IMG	fdiDRAFT24420
PebA1CROWA	<i>Crocospaera watsonii</i> WH 8501	DOE-IMG	CwatDRAFT_2853
PebA2CROWA	<i>Crocospaera watsonii</i> WH 8501	DOE-IMG	CwatDRAFT_0676
PebANOSPU	<i>Nostoc punctiforme</i> PCC 73102	DOE-IMG	Npun_R0730
PebATRICH	<i>Trichodesmium erythraeum</i> IMS101	DOE-IMG	Tery_5051
PebAPROMA	<i>Prochlorococcus marinus</i> CCMP 1375	DOE-IMG	Pro1749
PebAPROMM	<i>Prochlorococcus marinus</i> MIT 9313	DOE-IMG	PMT1686
PebASYNPY	<i>Synechococcus</i> sp. WH 8020	DOE-IMG	Ga0078671_111491
PebASYNPX	<i>Prochlorococcus</i> sp. WH8102	DOE-IMG	SYNW2020
PebASYN9605	<i>Prochlorococcus</i> sp. CC9605	DOE-IMG	Syncc9605_0423
PebASYN9902	<i>Prochlorococcus</i> sp. CC9902	DOE-IMG	Syncc9902_1906
Gasu_55890	<i>Galdieria sulphuraria</i>	Genbank (genomic)	Gasu_55890
CCMP769_PEBA	<i>Rhodospirillum rubrum</i>	MMETSP0011	7168_1

	<i>marinus</i>		
CCMP1999_PEBA	unclassified sp.	MMETSP1475	17257_1
SAG1380-2_PEBA	<i>Porphyridium aerugineum</i>	MMETSP0313	1488_1
CCMP736_PEBA	<i>Rhodella maculata</i>	MMETSP transcriptome	6728_1
CCMP3124_PEBA	<i>Erythrolobus australicus</i>	MMETSP1353	532_1
CCMP3234_PEBA	<i>Madagascaria erythrocladiodes</i>	MMETSP1450	49920_1
RCC2596_PUBS	<i>Ostreococcus mediterraneus</i>	MMETSP0932	4217_1
BCC118000_PUBS	<i>Ostreococcus lucimarinus</i>	MMETSP0939	5045_1
CCMP1194_PUBS	<i>Prasinococcus capsulatus</i>	MMETSP0941	2435_1
SL-175_PUBS	<i>Mantoniella antarctica</i>	MMETSP1106	8279_1
Ppat_PUBS	<i>Physcomitrella patens</i>	Genbank (genomic)	AFK50337
ZRMT-2001364	<i>Mougeotia</i> sp.	OneKP	ZRMT-2001364
STKJ-2007533	<i>Cosmarium ochthodes</i>	OneKP	STKJ-2007533
Penium_PUBS	<i>Penium margaritaceum</i>	Genbank (EST)	HO601495
kfl00619_0010	<i>Klebsormidium flaccidum</i>	Genbank (genomic)	GAQ90251
Gasu_58250	<i>Galdieria sulphuraria</i>	Genbank (genomic)	Gasu_58250
CCMP736_PEBB	<i>Rhodella maculata</i>	MMETSP	23469_1

		transcriptome	
SAG1380-2_PEBB	<i>Porphyridium aerugineum</i>	MMETSP0313	2204_1
CCMP3124_PEBB	<i>Erythrolobus australicus</i>	MMETSP1353	9977_1
CCMP769_PEBB	<i>Rhodorus marinus</i>	MMETSP0011	7399_1
CCMP3234_PEBB	<i>Madagascaria erythrocladiodes</i>	MMETSP1450	12871_1
CCMP1999_PEBB	unclassified sp.	MMETSP1475	3608_1
PebBGLOVI	<i>Gloeobacter violaceus</i> PCC 7421	DOE-IMG	gvip180
PebBFREDI	<i>Tolypothrix</i> sp. PCC 7601	DOE-IMG	fdiDRAFT24430
PebBCROWA	<i>Crocospaera watsonii</i> WH 8501	DOE-IMG	CwatDRAFT_5247
PebBNOSPU	<i>Nostoc punctiforme</i> PCC 73102	DOE-IMG	Npun_R0729
PebBTRICH	<i>Trichodesmium erythraeum</i> IMS101	DOE-IMG	Tery_5050
PebBSYNPY	<i>Synechococcus</i> sp. WH 8020	DOE-IMG	Ga0078671_111492
PebBSYNPX	<i>Prochlorococcus</i> sp. WH8102	DOE-IMG	SYNW2021
PebBSYN9605	<i>Prochlorococcus</i> sp. CC9605	DOE-IMG	Sync9605_0422
PebBSYN9902	<i>Prochlorococcus</i> sp. CC9902	DOE-IMG	Sync9902_1907
PebBPROMM	<i>Prochlorococcus marinus</i> MIT 9313	DOE-IMG	PMT1687

PebBPROMA	<i>Prochlorococcus marinus</i> CCMP 1375	DOE-IMG	Pro1748
HY2ARATH	<i>Arabidopsis thaliana</i>	Genbank (genomic)	NP_566344
HY2ORYSA	<i>Oryza sativa</i>	Genbank (genomic)	BAD87875
ELM1ZEAMA	<i>Zea mays</i>	Genbank (genomic)	NP_001105256
HY2LYCES	<i>Solanum lycopersicum</i>	Genbank (genomic)	BAD93478
Acv_HY2	<i>Adiantum capillus-veneris</i>	Genbank (genomic)	BAF02518
Pp_HY2	<i>Physcomitrella patens</i>	Genbank (genomic)	BAF02520
Chaeto_HY2	<i>Chaetosphaeridium globosum</i>	Genbank (EST)	HO384609
kfl00177_0010	<i>Klebsormidium flaccidum</i>	Genbank (genomic)	GAQ84116
AZZW-2019954	<i>Chlorokybus atmophyticus</i>	OneKP	AZZW-2019954
JPYU-2037290	<i>Marchantia polymorpha</i>	OneKP	JPYU-2037290
TCBC-2082726	<i>Nothoceros vincentianus</i>	OneKP	TCBC-2082726
FFPD-2057880	<i>Ceratodon purpureus</i>	OneKP	FFPD-2057880
XZUY-2012311	<i>Cycas micholitzii</i>	OneKP	XZUY-2012311
BYQM-2011005	<i>Posidonia australis</i>	OneKP	BYQM-2011005
FFGR-2000050	<i>Netrium digitus</i>	OneKP	FFGR-2000050

WDCW-2047152	<i>Mesotaenium endlicherianum</i>	OneKP	WDCW-2047152
BFIK-2028189	<i>Entransia fimbriata</i>	OneKP	BFIK-2028189
FQLP-2031011	<i>Klebsormidium subtile</i>	OneKP	FQLP-2031011
RCCRARATH	<i>Arabidopsis thaliana</i>	Genbank (genomic)	NP_195417
DZTK-2041775	<i>Batis maritima</i>	OneKP	DZTK-2041775
HYZL-2141617	<i>Akania lucens</i>	OneKP	HYZL-2141617

¹Sequences found in individual MMETSP transcriptomes are designated by the data set (e.g. MMETSP1475) and the sequence identifier within that set (e.g. 3608_1). Sequences found within the complete MMETSP assembly for a given organism are designated by ‘MMETSP transcriptome’ and the sequence identifier within that assembly (e.g. 23469_1). Sequences found in the Genbank database of expressed sequence tags are designated by ‘Genbank (EST)’ and the accession code for that EST (e.g. HO384609). Sequences derived from genomic sequences and found within the Genbank protein database are designated by ‘Genbank (genomic)’ and the protein accession code (e.g. NP_195417). Sequences found in the publicly available OneKP transcriptomes are designated by ‘OneKP’ and the sequence identifier for that transcript (e.g. XZUY-2049330). Sequences found in the DOE-IMG database are designated by ‘DOE-IMG’ and the locus tag for that sequence in the DOE-IMG database (e.g. Pro1748). The PCYA sequence for *Cyanophora paradoxa* CCMP 329 was inferred from a combination of genomic and EST data (Fig. S1a).