

Supplementary Materials for

Assembly of the algal CO₂-fixing organelle, the pyrenoid, is guided by a Rubisco-binding motif

Moritz T. Meyer, Alan K. Itakura, Weronika Patena, Lianyong Wang, Shan He, Tom Emrich-Mills, Chun S. Lau, Gary Yates, Luke C. M. Mackinder, Martin C. Jonikas*

*Corresponding author. Email: mjonikas@princeton.edu

Published 11 November 2020, *Sci. Adv.* **6**, eabd2408 (2020)
DOI: [10.1126/sciadv.abd2408](https://doi.org/10.1126/sciadv.abd2408)

The PDF file includes:

Figs. S1 to S9
Legends for tables S1 and S2
References

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/6/46/eabd2408/DC1)

Tables S1 and S2

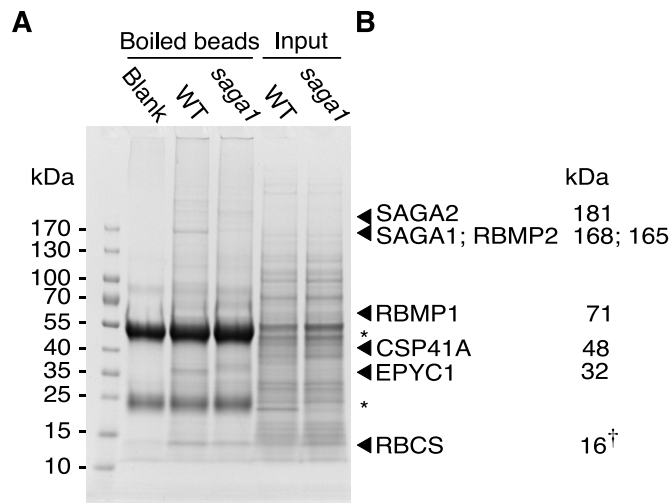


Fig. S1. A polyclonal antibody raised against the pyrenoid protein SAGA1 interacts with at least five other pyrenoid proteins.

(A) Coomassie-stained gel from the anti-SAGA1 immunoprecipitation experiment. First three lanes (identical to Fig. 1D): boiled beads alone (Blank), boiled beads after immunoprecipitation from wild-type (WT) and *saga1* lysates, respectively. The small and large chains of immunoglobulins used in the assay are labelled (*). Last two lanes: cell lysate input into immunoprecipitation experiment. (B) Calculated molecular weight (MW) of immunoprecipitated pyrenoid proteins based on the full-length sequence (46), except for RBCS (†) for which the sequence of the mature protein is known. The MW was not adjusted for predicted chloroplast transit peptides.

A

SAGA1 192 SDPPAP PPHSSSTAA AAAAATDVLFLVGS CP E L G E M D P G R A I K L A A V A G --- G G W A
 SAGA2 P P P P --- G S P E P V P V L T G S A A E L G A W E P S Q G V P L R R D P G A A --- G C E T
 At2g40840 Q D D S V V V Q F --- K I C C P D I C E G T S V Y V I G T I P E K L G N W K V E N G L R I N Y V D D --- S I W E
 1KUL T P T A V A V T F --- D L - T A T T T Y G E N I Y L V G S I S Q L G D W E T S D G I A L S A D K Y T S S D P L W Y
 1CYG T N D Q V S V R F --- V V N N A T T N L G Q N I Y I V G N V Y E L G N W D T S K A I G P M E N Q V V Y S Y P T W Y

SAGA1 246 A E A R L E L E S E - V A A K L L I M --- R D G T R M E W E L G P N R V L R G A L T A A A A Q P G T G A P P P R A
 SAGA2 A R A A L R M D R Q P L N A K L V L V A A G A G A A G T T A V W E E G A D R V L G H A T N --- P A A G P D R M
 At2g40840 A D C L I P K A D F P I K Y R Y C K V --- Q K E D S I G F E S G G N R E L S L H S I C S K Q E Y I --- V M S D G
 1KUL V T V I L P A - G E S E E Y K F T R I --- E S D D E V E W E S D P N R E Y T V P Q A C G T --- S T
 1CYG I D V S V P E - G K I I E F K F I K K --- D S Q G N V T W E S G S N H V Y T T P T N T --- T

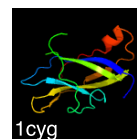
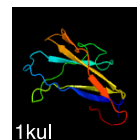
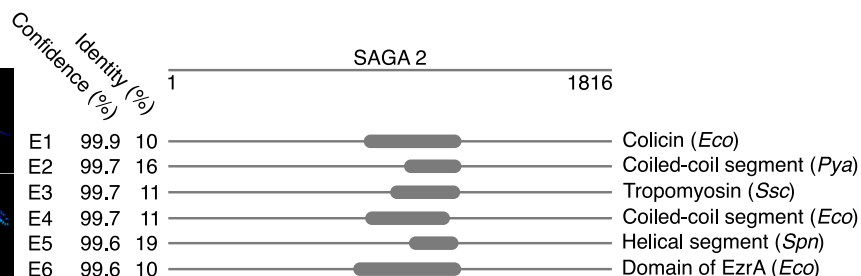
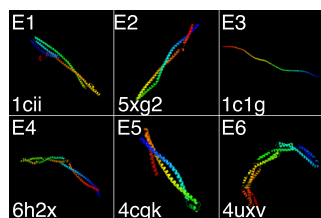
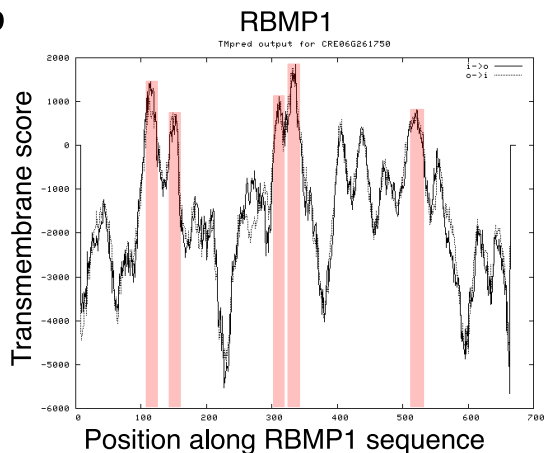
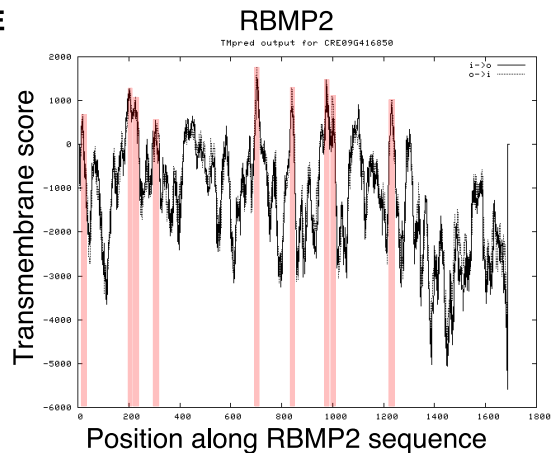
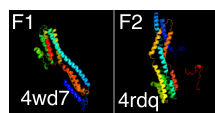
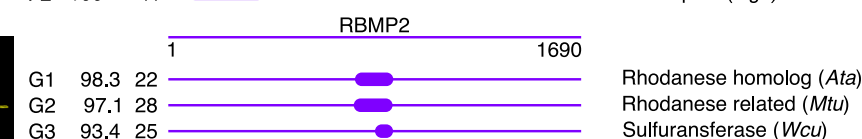
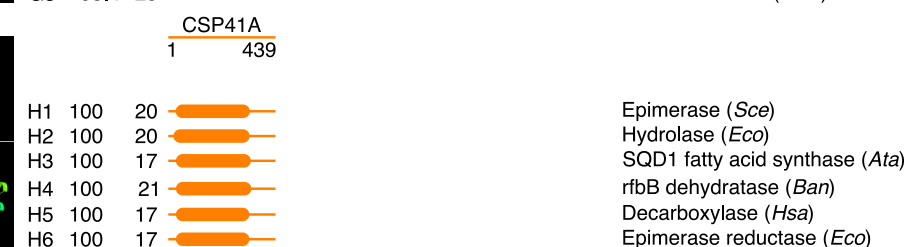
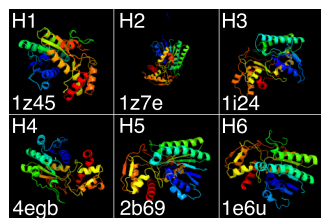
B**C****D****E****F****G****H**

Fig. S2. Predicted functional annotation of pyrenoid proteins immunoprecipitated by the anti-SAGA1 antibody. (A) SAGA1 and SAGA2 contain a predicted starch binding domain (belonging to the ubiquitous CBM20 family). Conserved residues predicted to be involved in starch recognition are highlighted (magenta). Numbering is for SAGA1. The partial alignment around the starch binding domain includes examples spanning the entire tree of life: a land plant (*Arabidopsis thaliana*, At2g40840), a fungus (*Aspergillus niger*, 1kul) and a Gram+ bacterium (*Geobacillus stearothermophilus*, 1cyg). (B) Crystal structures of two starch binding domains from (A). (C) Predicted regions of structural similarity between SAGA2 and experimentally determined protein structure data deposited in the Protein Data Bank (47). Six predicted structures are shown alongside PDB template ID, confidence and identity percentages. The position of a predicted functional domain is shown along the protein length of SAGA2 (to scale). Abbreviations: *Eco* (*Escherichia coli*), *Pya* (*Pyrococcus yayanosii*), *Ssc* (*Sus scrofa*), *Spn* (*Streptococcus pneumoniae*). (D) Predicted transmembrane domains in RBMP1 (48). Bestrophins are calcium-activated ion channels that assemble as homo-tetramers or homo-pentamers (29, 49). (E) Predicted transmembrane domains in RBMP2 (48). (F to H) Predicted regions of structural similarity between RBMP1 (F), RBMP2 (G), and CSP41A (H) with data deposited in PDB, as for (C), above. Abbreviations: *Kpn* (*Klebsiella pneumoniae*), *Gga* (*Gallus gallus*), *Ath* (*Arabidopsis thaliana*), *Mtu* (*Mycobacterium tuberculosis*), *Wsu* (*Wolinella succinogenes*), *Sce* (*Saccharomyces cerevisiae*), *Ban* (*Bacillus anthracis*), *Hsa* (*Homo sapiens*).

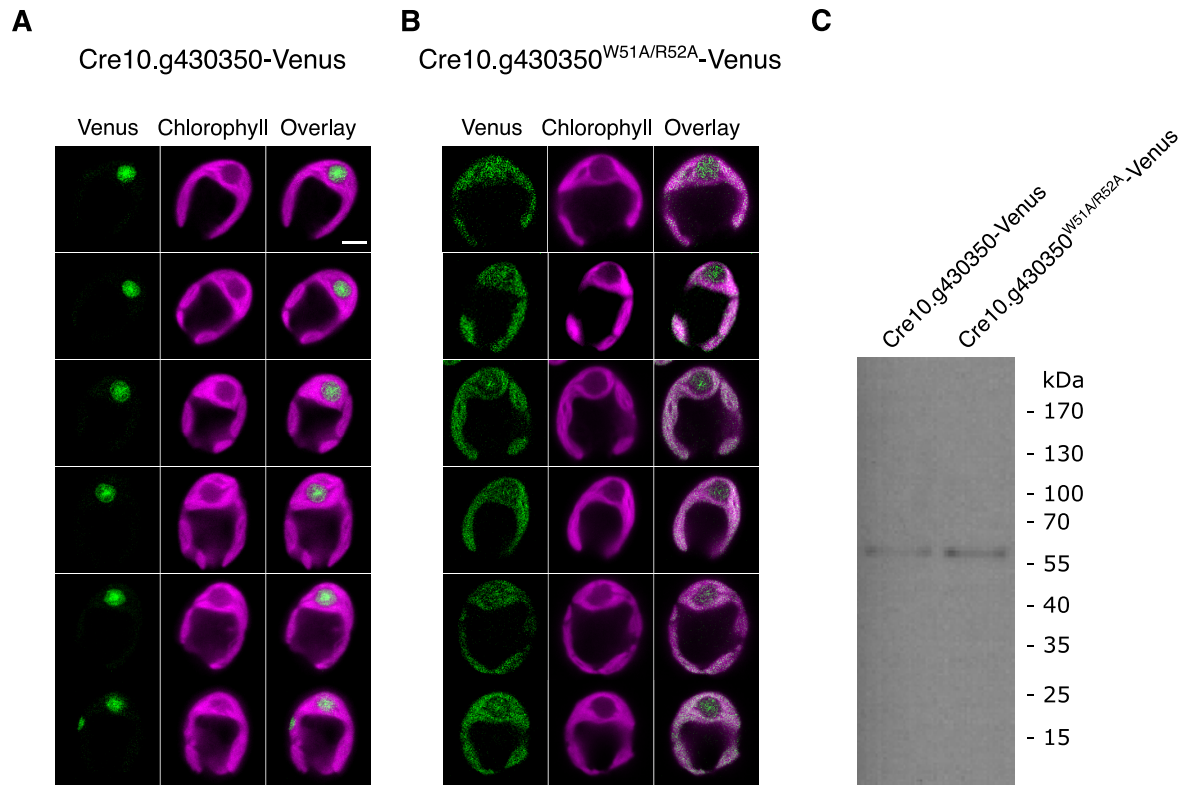


Fig. S3. Subcellular localization of Cre10.g430350. (A) Supplementary confocal images of the native localization of Venus-tagged Cre10.g430350. Scale bar, 2 μ m. (B) Supplementary confocal images of Venus-tagged Cre10.g430350 with mutations W51A and R52A. (C) Anti-FLAG western blot validation of full-length expression of proteins tagged with Venus-3xFLAG.

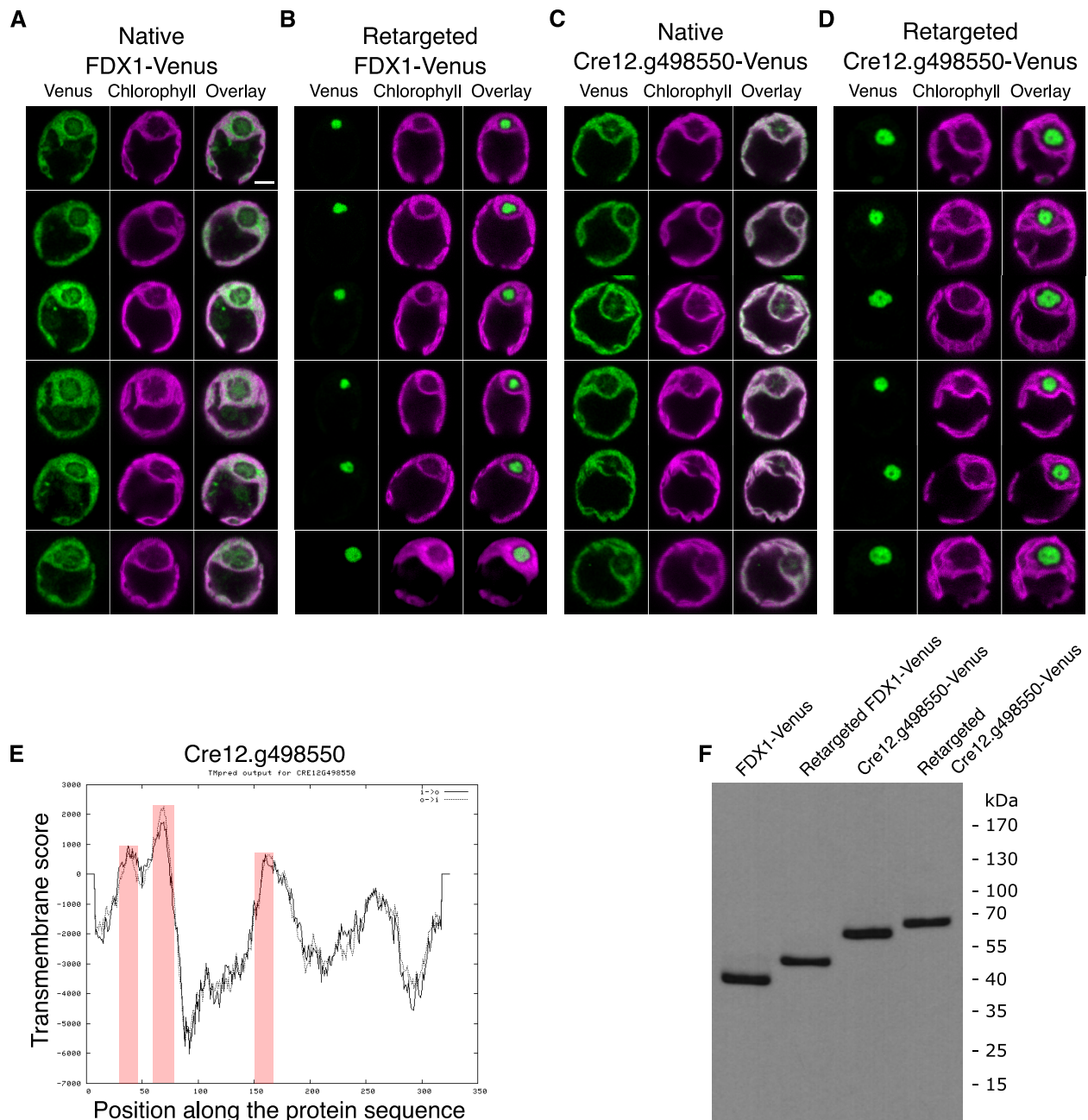


Fig. S4. Subcellular localization of native and pyrenoid-retargeted chloroplast proteins. (A) Supplementary confocal images of the native localization of Venus-tagged ferredoxin 1 protein (FDX1). Scale bar, 2 μ m. (B) Supplementary confocal images of FDX1-Venus with the C-terminal addition of three copies of the C-terminal SAGA2 motif. (C) Confocal images of the native localization of Venus-tagged Cre12.g498550, an uncharacterized chloroplast protein homologous to a Viridiplantae conserved magnesium protoporphyrin methyl-transferase involved in

tetrapyrrole metabolism. **(D)** Confocal images of Cre12.g498550-Venus with the C-terminal addition of three copies of the C-terminal SAGA2 motif. **(E)** Predicted transmembrane domains of Cre12.g498550 (48). **(F)** Anti-FLAG western blot validation of full-length expression of proteins tagged with Venus-3xFLAG.

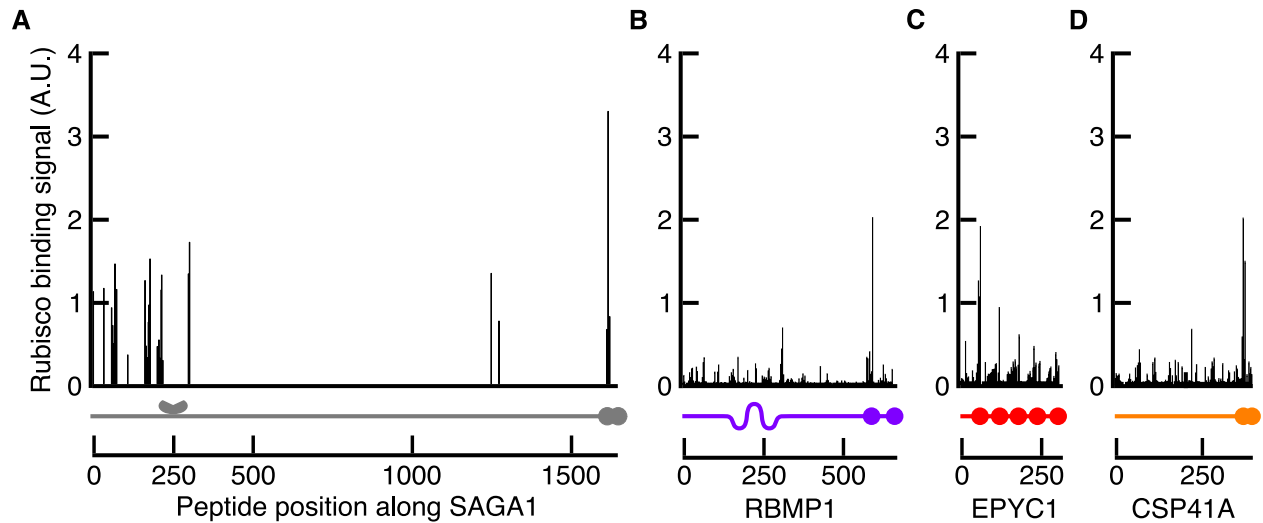


Fig. S5. Rubisco-binding measured by peptide array. (A) Array of 18-amino-acid peptides tiling across the sequence of SAGA1, (B) RBMP1, (C) EPYC1, and (D) CSP41A. Arrays were synthesized and probed with Rubisco. Binding signal is normalized to a control EPYC1 peptide (same as for Fig. 3B and 3C, corresponding to one unit of binding). The positions of the predicted motifs are indicated to scale below each graph. The peptide corresponding to the C-terminus does not accurately represent the Rubisco-binding motif in this assay, as the peptides are linked to the cellulose via their C-termini, eliminating the carboxyl group which appears to be important for binding to Rubisco (based on the observation that internal instances of the motif are typically followed by an aspartic or glutamic acid, each of which carries a carboxyl group). The binding response was quantified by summing all the pixel intensities in a constant circular region centered on each dot on the peptide array.

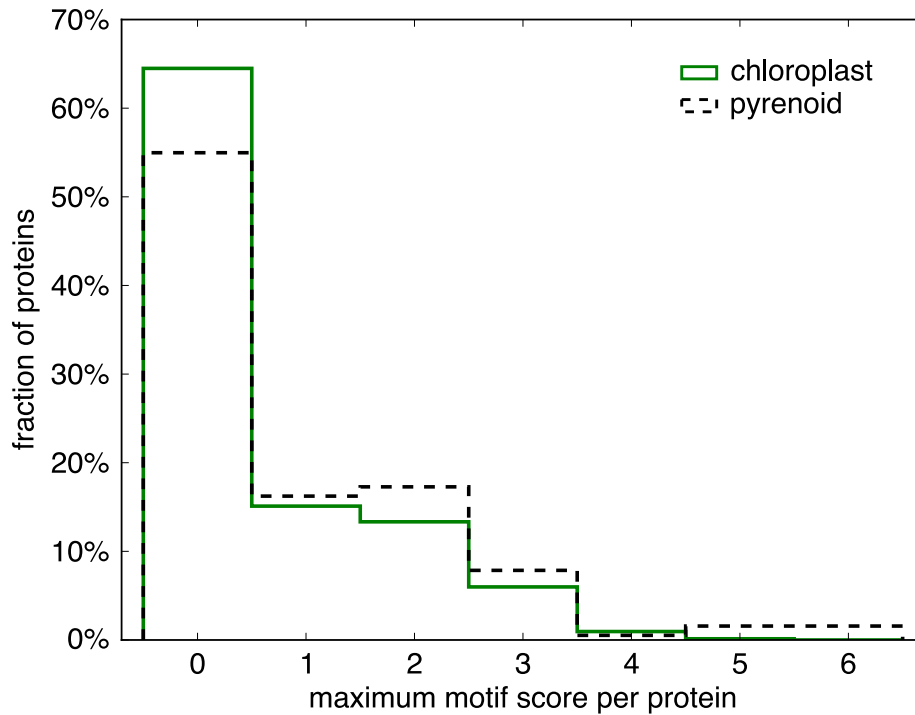


Fig. S6. High motif scores were modestly enriched among pyrenoid proteome proteins relative to predicted chloroplast-targeted proteins. The fraction of proteins with each motif score is shown. Pyrenoid: proteins found in the pyrenoid proteome (17). Chloroplast: proteins found in the chloroplast proteome (50) excluding proteins found in the pyrenoid proteome (see Supplementary Table S2). The pyrenoid proteome has a distribution of motif scores that is shifted toward higher scores ($p = 0.047$ after excluding the six proteins used to define the motif [EPYC1, SAGA1, SAGA2, RBMP1, RBMP2, CSP41A]; $p = 0.0066$ otherwise).

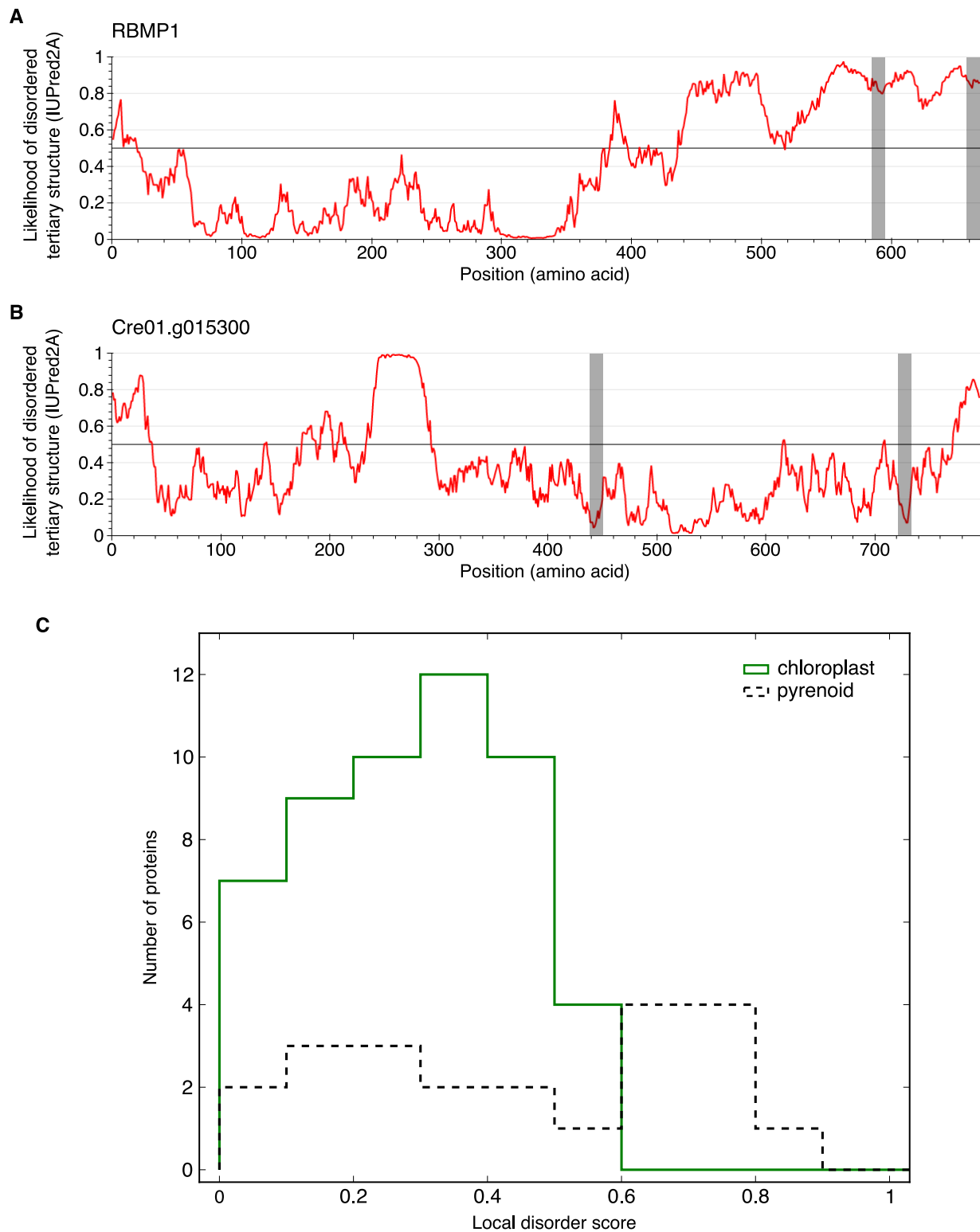


Fig. S7. The motifs of pyrenoid-localized proteins are more frequently situated in disordered regions compared with the motifs of other chloroplast proteins. (A) The likelihood of

disordered tertiary structure as predicted by IUPred2 (51) was plotted as a function of amino acid position along the length of the pyrenoid-localized protein RBMP1. Each grey bar highlights an 11-amino acid region centered on the tryptophan of a Rubisco-binding motif, in which the plotted values were averaged to obtain a “local disorder score”, reflecting the likelihood that this motif is in a disordered region. **(B)** For illustration purposes, the same analysis is shown for the non-pyrenoid chloroplast protein Cre01.g015300. **(C)** The analysis was repeated for proteins found in the pyrenoid proteome (17) and/or the chloroplast proteome (44). Only motifs with scores of 3 or greater (relatively high similarity to validated Rubisco-binding sequences) were considered. When multiple such motifs were found in the same protein, their average score was used for that protein. The number of proteins with each local disorder score is shown, for either the proteins found in the pyrenoid proteome (“pyrenoid”) or proteins found in the chloroplast proteome excluding proteins found in the pyrenoid proteome (“chloroplast”) (scores are shown in Supplementary Table S2). Proteins in the pyrenoid proteome have a bimodal distribution of motif-localization disorder scores, with a peak shifted towards higher scores, and the distributions have different means (Welch’s t-test $p = 0.007$).

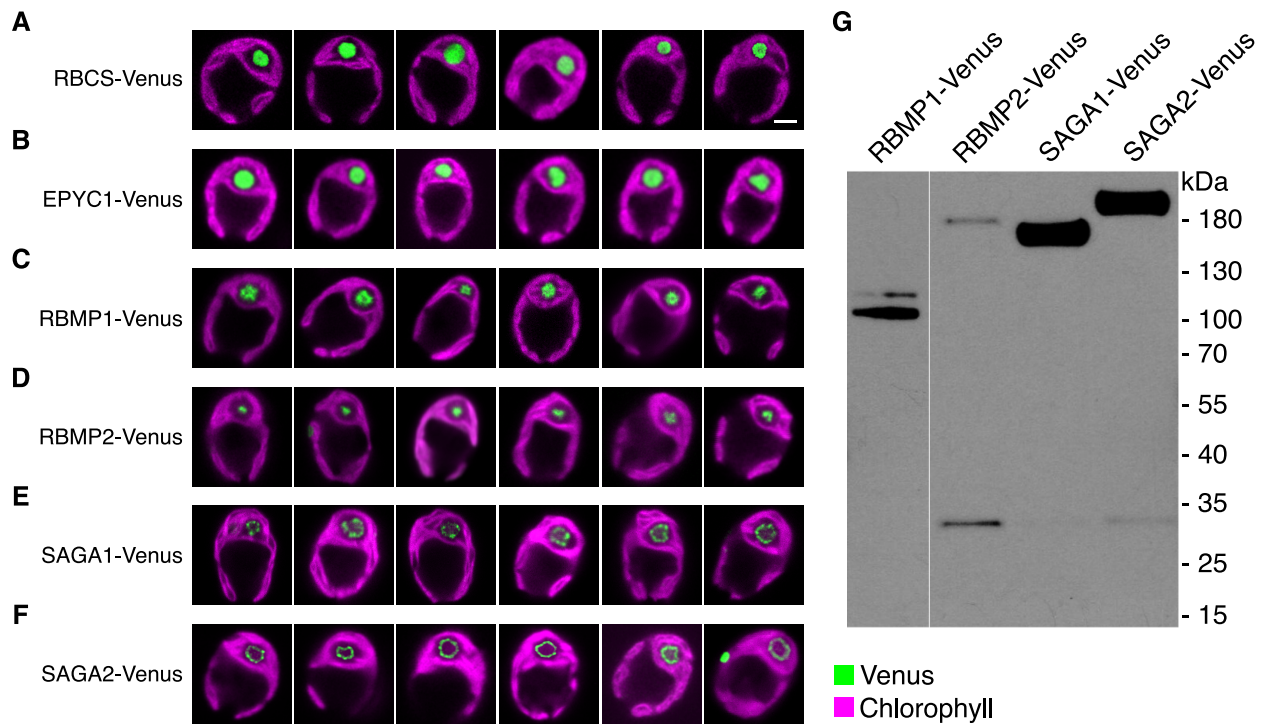


Fig. S8. Subcellular localization of native pyrenoid proteins. (A) to (F), Supplementary confocal images of the native localization of Venus-tagged RBMP1, RBMP2, SAGA1, SAGA2, EPYC1, and RBCS1, respectively. Chlorophyll autofluorescence delimits the chloroplast. Scale bar, 2 μ m. (G) Anti-FLAG western blot validation of full-length expression of proteins tagged with Venus-3xFLAG: RBMP1 ~99kDa, RBMP2 ~199kDa, SAGA1 ~191kDa, SAGA2 ~212kDa. For RBCS1-Venus and EPYC1-Venus, see (18).

A

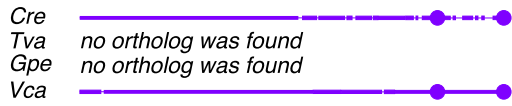
SAGA1 N-  -C



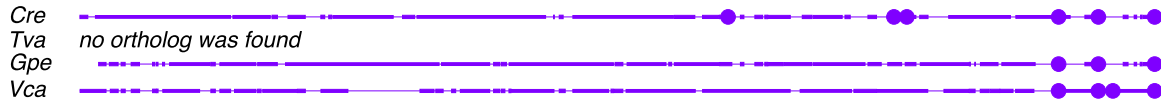
SAGA2 N-  -C



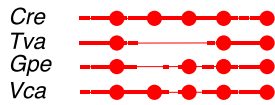
RBMP1 N-  -C



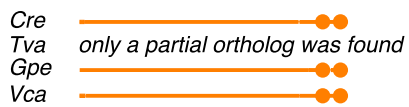
RBMP2 N-  -C



EPYC1 N-  -C



CSP41A N-  -C



B

SAGA1 (2/4)

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

AAALAA-----DGELQRLTAQLQSVSAQLETTMTEMASARAAVVAKERAMAELEERLA
PPPRGDGASPSADAGELVRLTSQLQSVSSQLDLTLAEMAAARAAVVAKERAVAELEERLA
AVPAGGGS-----ADEVARLASQLQAVSSQLDRTAADMAAARAAVVAKERAVAELEERLA
GAAAAS-AVPTADPSEIQRLASQLQNVNAQLELTVAEMAAARAABAERAVAEEMEERLN
. . . *: **:*** *.**: * :**:*****.*****:***:***

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

AAAARDAERKQELQAVRSQQLADAVARTSAAADASATADTLASKTAAALQSELAATKQMYE
TAAARDAERKQELQALRTOLESALVRSSAVADANATADSLASATTSALQOGELQOTKQMYE
AAAARDSEERRQELALRAQLTEAQRSSAVADASVTADSLASATAAALQAEMLQTKQMYE
RATAKDAERKQELQALRAQLADALARSNAAADASATAEGLASATAQALQVELQOTKQLYE
*: **:*** *:**:*. * .*:*.***.**: *** *: *** ** *:**:**

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

STLAEALAEANSINANSRALLSQLEENFGAASSQYEATLQLFDSRVTELTRQLEAAERARS
ATLKEQLDANTINTNSRALLTQLEENFSAASTQYEATLELFDARVTELGRLQLEVAESTRS
ASLAEELTEANAINANSRALLAQLEDNFSAASTQYEATLELFDARVSELTRQLEAAERARS
ATLKEQLDANTMNANSRLLTQLEQNFSAASSQYEATLELFDARVSELTRQLEAAERSRV
*: * * .**:*.**:***:***:***:***:***:***:***:***:***:***:***:***:***:***

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

EEVSALEAALASTRTGQQEVLQAIQENFMAASQQYETTLTLYDTRVVELEAALEGAESRR
EDVAALEAALESTRSGQQEILSQIQENFVAASQQYETTLTLYDTRVGELEEALEVAEQR
DEVAGLEAALASRQQQEVLTAIQDNFVAASQQYETTLTLYDTRVVELEAALEAAEAAR
EEVRSLEEALSTRSSQQEVLNQIQENFVAASQQYETTLTLYDTRVVELEAALAEAESRR
*: * ** *: * ***: * ***:***:***:***:***:***:***:***:***:***:***:***:***

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

LEDVSLLEGELEEARQAQATLLGELEESLTATSQYEATLKLKLYEDKIDELNAQLEAVQAD
SSDVTALEAALDAARAGQAALLGELEESLTATSQQYETTLKLYEDKIEELNVQIETVQAD
SEEVGLEEALEAARSQAALLGELEESLTATSQQYEATLKLKLYEDKIEELNGQVEAAQAO
SEEVSVLEAALDAARQAQASLLGELEESLTATSQQYEATLKLKLYEDKIEELNTQMESLQMD
.**: ** .*:*** .**:***** ***** *:***:***:***:***:***:***:***:***

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

YSSRIGELNTNISTEQASQRKAAEFGAQLADLTQQQLRGAAREEQTLQRLARVEAEAAE
YDKRITSLDDGMAQEQANQRKAAEQYAAQLTDLQDKMRAADREAQLSERLAVVEAEATE
YDGRTAELNNRIGEEQAQRRAEQYSAQLSELQAQLREATEREAEQLGRLEERQAEAAE
YGSRIEQLNSDIAQQANQRQATEEYSAQLAQLQQLREATEREVQLAERITRTETEAAE
* . * .*: .: .**: ***:***:***:***:***:***:***:***:***:***:***:***:***

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

ARASLQSELDLRRDGSERQSALQQQLAQVQ-----DTATQFEQLARDQAASLANAQE
ARTRLEGELESIRNEAKQRQAELEGGQLAQIQANHGGRADTASQFEQLARDQAASLSTAQE
A-----
RETRLAEQISRTEADAERQMQLAERLSRTEA-----EAAEREMQLAEQLNRTEAEAVE

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

RQGELVAQVLALQQETAAREAEEL-----RSTTARYEELAREQAAGLATSQERQAEELVA
RQGELVAQVLALQQETAARETEL-----RDATLKFEELAREQTAGLSTAQERQNELVA
-----RDTSSQYESRAREQSAALASAQERQAEVVA
ARVRLEAELEGIRREAERQGELEGGQLAQVQSTASQFEELARDQAASLATAQERQAEELVA
. .: .:*. **:***.**:***:***:***:***:***:***:***:***:***:***:***

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

QLLALQQESQAKESELRAALAAERQANEKGSILLRERGAALDSSLAAQORLQMELELQD
QLLALQQEAQAKEGELRGLLQGERQAKEKSAILLRERVAALESALAAQORLQMELELQD
QLLALQQESQAKEAELR-----D
QVLALQQETAAREAEELRGLLQSERQAREKGAIVLRERVAALDAALATQORLQMELELQD
*:***:***: * . * .*** :

B

SAGA1 (3/4)

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

SLGSREAAAAQAKAQVLEAAAAEAALTTQLKEAREAAEKRESALRKEVDVLRNATSEAS
SVGAREAAAAQARAQVMEAAARTNEEQLRGQLKATRDEAEDREAALARELDELRLSATVTE
QVASREAAAAQARAQVMEAAARTNEEQLRGQLKATRDEAEDREAALARELDELRLSATVTE
ELASREAAAAEKARAEMSEAAAREVEQRLTGQLKATREEAETREAALQREVDALRLQATVAE
.:.:*****:*.*: *** * * :** *: ** ***: .:* ** :. :.

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

AKLKEQKSQSEELAAAALAAQARAARELEEARAAAAEAAAAAAAAEAAKAKERELGELA
AKLREQVESTREAGSLAAQARAARDLEESRSAAAAARQQLSALQQAASAKERELSELL
AKLREQVASSQAEASSLAEQARAAREVEEARAAAGAARQQLAALQTAGAAKEAELGKLL
AKLQEEMAAQEQAAASLAEQSRAREVEEARAAAAAVRKQLAALQQAAGAAKEAELGEML
.: : *.:** **:*::***:***. .* :* *. *** **.::

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

DSARAEAEERVAAKLAEAEAAAASATGRERELGEQLAGKEEALTAARAEVAAARRQLEEAQ
ASSRSESEQLSRVAEQEAALQAAAAKERELLEQVGOEAALTAARAEISKQSQSLSEAQ
AAAREESGRLQSLGERESALAAAAAKEELGERVRRAEALAAAQALGALRGELAAASR
RSARTDAERLQSLSEREVALKAAQERERQLSEQVGRTEELAAAKAEVGLERNQLSEAQ
:* :. :. :.* * * :* . * : * * : * **:*:*:*.. . : * :.

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

QQLADRDAQLADLPDLQRLGLMDELSAELAEAEHAAAQERLAAAGAEAEAAQALRA
QQVLERDQGLSTMSAIESRLGLIDELNAELSSLMEQYNSARAEALGTVRSSNDTAQAQLAS
Q-----QARLRS
QQLADRDTQLAAVSAIESRLGLIDELNAELSSLMEQYNAAKEELSTVRTSYDKAQAQLTS
* ** * :

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

VQADYSAAQSKLVELGRSYEALKAEGEQNAQWDAEEELNEIAELQEAVVKYRETISD
MTSEYESARAQLAKATQYEDVKEESVEQARWEAAEETYLSEIAELSEAVVKYRDTIAD
LSSDYEAARKQLASVSSSQAEAKAKMDEVSEWEAAEGYLSIEIAELSEAVLQYRETIGQ
MSSDYGAARKQLAVVSYLLEEARKKAEQ SARWS TEEAYLADISELSEAVVQYRETISD
: ::* :*. :*. . : ** : *.:** * :*:**.*:*:***:*:..

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

LNADLAEINQRYSSLVQDRQR-ATEDGGKRSEELGAAQSRQAQLQEQLKRSESRLERLAA
LNSDLAALMEQYDTLLRDRQR-ASDDRGQRSEEAQA-----LRDQLKRSEGRAERLMT
LNAELASANQRVSSLLQDRQRSASEEGGRSAETDALAARSAAAEKLRSELRVDRLMA
LNNDLAGLMDRYNNLLRDRQR-AGDDSSRRSSEMEALNNKVVLQAEALKRSEGRYEQLLG
** :** :. .:.*:*:* * :. :.* * * * * : * :*** * :.*

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

QLADEQEARRRDAEVAAAQREALAAEAKKAAAASAAASKGAGAGAEAAAQQVAELQVALQ
ALADEQEARRREAKAAEAQQAAMMVELRQAATSSSTRGGRD-----GPDVSQLQIQLO
QLAEEQAARRREAQAAEALRSELAK-----
QLADEQEARRRDVEAAEAQRQALMRSVEEASTAGYSSS-----AAAVSELKIQLO
:* * *:..* * . :

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

QSEAVRLAQVSELQERLAKANQAAAQMRGQQSGESPRGGAGAAAAGPEAARLRAELEGRM
QSEA-----VTELQSRSLRADQAVSOLQAQLTSDSPR-SSGSGFGGADATRVRQYEQYQRI
-----VTELQERLARANAVAALQRQQQQQQV-AAPAGGGGGAARLLSEYQTRM
QSEAVRIAQVSELQERLARANQTLSQLSANSEAAASRQ-----AGGAPSARVRQYEQYQRI
::*.*:*.: : : : . * :*: * : *

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

AAMVREYGRSTASRAALRTLRLQSLTLIRGGVAAQSAQERLMDELRAAEMEMDAMSALG
AAMTREYTSKTSRRNALRTRVRSLLQIRGGVAAQQAQERLLEELRAAEGYMDMAALP
AEVAKDVS SSKAAGRNALRSVRQSLQLIRKGGGAAAAQERLMEELRAEAQLDAIASLS
ATMVKDYSTKSLGRNALRAVRQSLLLIKRGGGAAQQAQERLMDELRAAEDDLELISLSLA
* :. :. :.* ***:*** * ..*:*.* ***:*** ***:*** :. :.*

B

SAGA1 (4/4)

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

TKYA-LPAPASPRGAIEAASSAAAADAAATARDDAAAAAQRRPAANPAAAEYEDAGSN
LKFAAQSPPPAAATPPPATPTPPPVESPTSGAKGGGS-----TSSDRGGSSSVA
AKAAPLPPQPAAT---GDASPPAWSSSAKGLPAPGA-----PPGSSLPSPAFT
AKYLQEQQQQQLPGLSADRSRGLEPSSSDAAPGAGRSSNSRSSNSSSSSSSSTPGNG
* . .: .:

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

KDNLFTRSRTSTGSSSSSSGSGSSGSSGGGGGTSKAQPQPGRST-----SA--DWR
VEGMFSSRSRVNKDSS-----SGGSSSSGGGRPALV-----RSN-----SPS-DWR
RTGMPPSPPPVSPSS-----SGG---GGSGPPASAPQRSPSGRSN-----SPPLDWR
GGGQPQYQSIFERS-----VGGGGGGGGGGGSAPSRSSKSSTVTPSTPPPPSDWR
. * ** .*.* *. .. ***

Cre Cre11.g467712.t1.1
Tva PNH05753.1
Gpe KXZ43034.1
Vca Vocar.0009s0363.1.p

RLVSG-----GDAAGKDADGGGRGTGDSPTRRAFGDWRKNI
KQITGGDGGKGGKGDAGGGGQDDDRGAGASPTRRNFGDWRAM-
KQLGG-----
QRLGA-----GEGGDSGAARGVSPTRRITYGDWRAG-
. : .

C

SAGA2 (2/3)

Cre Cre09.g394621.t1.1 GRVSKHSGAGGSASDSEGETVTKANWREALAAAHDAGSDGE-----
 Gpe KXZ45943.1 -----
 Vca XP_002951554.1 TSPSSSSSGSAAAAAKRSSPRTATSSNWRQMLEEEEEAEHREELQQHQYHQQQQVEVEE

Cre Cre09.g394621.t1.1 -----EAAVARSSSPARAGAGWPRATSPGARATSPG---
 Gpe KXZ45943.1 -----
 Vca XP_002951554.1 LFGGFMLGSSRRGDVEGMVEEQSGIGRSLDSRAGSPARMGSLSPRGSFSPSPSRSPSPGA

Cre Cre09.g394621.t1.1 -----RSQPVAAALAAEHARMADLERQHMEAMEAARQAQDAQRQVESQLAARAA
 Gpe KXZ45943.1 -----
 Vca XP_002951554.1 NDAVQQAEEQQAAAAAAAAAAAAAAAAARLAALAEHLSAMEAARLAAVEAQKQLEGQLAAKAA

Cre Cre09.g394621.t1.1 LEAERVAEYESAMAALKAQVEAAQAEAAAANRLALETMGQQLEASRGREQDLSGELGLMQS
 Gpe KXZ45943.1 -----
 Vca XP_002951554.1 EEEARIAEYEAQMAMLRQQVDAQSEASNQLSLHALQQQLEASRGREADLSEELTAMQG

Cre Cre09.g394621.t1.1 RLAEAAARSHNSNSDAVNALESALQAARSQYGALAEMQRADAAGSYSLRLLERLEAAYTG
 Gpe KXZ45943.1 -----
 Vca XP_002951554.1 KLAEAARSHTSNTAEAVDVLESALQTARQQYESALTELQRADAAGSYSLGLLDRLEAAYLA

Cre Cre09.g394621.t1.1 AVDQYEDLEKELVAAKNAQAQSESMLAAMEAQVATSMKLAGSVQVYEKQVAVLQEQQLQMS
 Gpe KXZ45943.1 -----S
 Vca XP_002951554.1 AVDQYEEVVAELATAKASQQSQAMLQTLAQVAASVQLAGSVEVYEKQVQMLQEQQLQLS
 *

Cre Cre09.g394621.t1.1 AEAHEAELGALQGQLASVEARASQTEDLYESMLKEMKSALDDALAREQVLHRELSSMQSR
 Gpe KXZ45943.1 ASAHEEELSALSGQLRGLARSQAQTEELYESMLKEMKAALDDALGREAMLHSELSSGLQSR
 Vca XP_002951554.1 AQAHEAELSLLQGQLRSVEVRSQQTEDLYESMLGEMKSALDDALAREQTLHAELNSMQRR
 * .*** ** .*** .:*.*: **:* **:* **:* **:* **:* **:* **:* **:* **:*

Cre Cre09.g394621.t1.1 AEEFQRQVEKAAAGMGESTRQEEALRKEVARLRRVIEQYK-----EYLEEHDGELGRSLE
 Gpe KXZ45943.1 TEELQAQAAQAAAGVSESRAAEEAMRKETQRLRQVIQQYK-----EYLEEHDGELGRSLE
 Vca XP_002951554.1 ADEFQKQVEAAAAGMSDSARQEEALRKEVARLRRVIEQYKHPGSQEYLTEHDGELGRSLE
 :.*:* * .***.:.* **:* **:* **:* **:* **:* **:* **:* **:* **:*

Cre Cre09.g394621.t1.1 REEALDAQLSSLOAVVSQQESQLRGALERQQALSAELEEQRRTADEYQAKLEEQQRELAE
 Gpe KXZ45943.1 REEALDAQLTGLQVVAQQEAAALRAAERQEALAAELEAQRASAEAAQAKLEEQQKELAE
 Vca XP_002951554.1 REEVLDAQLTSLQAVVRQQEEQLQAALERQDELSRELEHQKTAEDYKSQLEQQAAALQE
 ..:** ** ** * ..* **:* *: ** * .:.* :.:*.*** * *

Cre Cre09.g394621.t1.1 AAAKQAAMNERADSMTREMGRLQELTESYKAKLEASQSEVEAAAAREKAAVAAARAAGER
 Gpe KXZ45943.1 AAAMQAGMRERAESMTAEISRLQELTEGYKTKLALTESEAADAAREAEALRKVEAAGAR
 Vca XP_002951554.1 AAEEKQAAMQERAEAMQREMDRLQGLTESYKTKLEASRSEALAAAREAEALVRQVEASGTR
 ** **.*.***.:* *:.*** **:* **:* : ** .*** : .:* ** *

Cre Cre09.g394621.t1.1 EGDLLGRVEKQLQANIASYKSKLQEQERSLSTAAQRESSLQAALAEQEVERVALAEQLHD-
 Gpe KXZ45943.1 ETDLLGRVEKQLQLSIAQYKSKLQDQAASLSAAEREGALTASLAEASSETAALKQELSSL
 Vca XP_002951554.1 EGDLLGRVEKQLQASIAQYKSKLQDQESSLSAAAAREGELMAALRDQEGDSADLRHRELEL
 * ***** .***.***.* **:* ** * . * : . : . * * .

C

SAGA2 (3/3)

Cre Cre09.g394621.t1.1 -----AVETITTSQKQVLENATAAQEEVL-VMAQQVQTL-----
Gpe KXZ45943.1 GGQYRKAVDTIASQQEQLEEAAKVVAKVD-GMAVQVQTL-----
Vca XP_002951554.1 GGEYRAAVETIEEQGRLEAVTNAAEQHEAAMAVQVQTLQPALHKSGTIRVRYIPPPPTR
:* .*: ** .: . : ** ***

Cre Cre09.g394621.t1.1 -----QVELSAERNRHEEMQQVQMALDEATTANKERLQVIGGSGG
Gpe KXZ45943.1 -----QVELSSERSRHNEELTAVAAALEEAAAGDRSRM-----
Vca XP_002951554.1 CAPSTFKTLKTCILKPHPSQVELSAERTRHTEEMQQVQQALEAATEVDRLRMQGRAQQG-
*****:*.*.*.*:*: * **:* *: .: . *:

Cre Cre09.g394621.t1.1 GLPGSTSASASGSASGSTSASGTTAALQPSASAGSSAKADSDLEFAADDAMASGGEIDYD
Gpe KXZ45943.1 -----
Vca XP_002951554.1 --RSRHRMRREGNFMVAVTMEGFQAGI-----

Cre Cre09.g394621.t1.1 QAEFEMQKAVLKAIEYEVKVTALKEALAVVKNALAPGAVKRPTAAAMPPPARAAAPAPAPP
Gpe KXZ45943.1 QAEFEMQRAVLKAIEYEVKLTALKE-----
Vca XP_002951554.1 RAEFEMQKAVLKAIEYEVKLTALKEALVVKNALNPATTKNA-----
.*:****.******.*:****

Cre Cre09.g394621.t1.1 AAARPAPPQAEERREQPPI PAPQPPQQAQPQAQVQPKAEKPKQEKPAAPAAAPPAPAPAPAK
Gpe KXZ45943.1 -----PRPPPSRFTLTEIVT-----
Vca XP_002951554.1 -----PTVAPPQQQQQQSLPP-----
** . .: .

Cre Cre09.g394621.t1.1 TKPDWREQAQAPVQAAAAAPVQAAAKPAEAAAAPATASTKPEAPKPAPAASSAPSSTTS
Gpe KXZ45943.1 -----
Vca XP_002951554.1 -----

Cre Cre09.g394621.t1.1 GADSRSPSPRTLAWREAAEAQEREQESKRQVEEQARIAAVAEAEARRAEDEFRRKHADDVP
Gpe KXZ45943.1 -ARSRSPVRQHLLA-----EHSARRAEN-----
Vca XP_002951554.1 -AQHRTASP-----
* *:. .

Cre Cre09.g394621.t1.1 AGGRGKTLREIINMTATERAAVVQETESLPRAASPARNASENARMRGWPVRIADMCPEDK
Gpe KXZ45943.1 -----HDH
Vca XP_002951554.1 -----

Cre Cre09.g394621.t1.1 EQQEQQEGGAKAERESERGGQAAGSGKRSESSERSESSERASPSKGGKGNRWANLANGES
Gpe KXZ45943.1 SSNGNGNGNGASANGNGHGEAATAAPSPSS-----SPGRSSRWASMA-----
Vca XP_002951554.1 -----VAAATAPLPAAASSSPSLTPT-----FATKVKWASMASELA
: *:. . * :. . .**:*

Cre Cre09.g394621.t1.1 AGAAASAASSSSSAAAAAEKQERQRGSGSNTSSSNGASSNGASSSRNGTHAAGEDVREV
Gpe KXZ45943.1 -----EEMAKVRGGGKGAADAAAGSGSAEEGSGASQRG-----
Vca XP_002951554.1 AARGEQPSASTSAAIAPVDANISVERNG---KVISHGVNNGKELSVAAGAEAQVD-----
*.. : . . . *

Cre Cre09.g394621.t1.1 GKDDAAEAGRSASGRRNASPVRR-TAIPANWRDAL
Gpe KXZ45943.1 -----SSRGGTSPTPRQR-TRVPANWRDAL
Vca XP_002951554.1 -----SVSGRDSSTGVRRKTSIPSNWRDAL
* * .: . * * :*:*****

D

RBMP1

Cre Cre06.g261750.t1.2
Vca Vocar.0002s0473.1.p

MQCQLKHGARPOQSQRPNWLPARAATLRPAVQHGVRRLTGLGVKAAAAPLEDKMPADMTT
MQSQLQPRQLQLQGTRLNWLQQRSCVQRRSLRVDATSG-----AAPPPPAGKELSNDMVT
.*: . * . * ** *:. * ::. .. * **.* .*: . **.*

Cre Cre06.g261750.t1.2
Vca Vocar.0002s0473.1.p

RQYRRVYDFALWAKHRDVNRYLYNLRTIPGSRIIRQLSQPMGVVLAWAALFGFYETCLE
RQYRRTVYDFSLWAKHRDVNRYLYNLKTIIPGSRIIRTLGQPMGIVLAWAAMFGFYETCLE
*****.***:*****.***** * .***:*****:*****

Cre Cre06.g261750.t1.2
Vca Vocar.0002s0473.1.p

AGVLPSYLPKMTLMSAEPQGLTSFALSLLLVFRTNSSYGRFDEARKIWGGILNRARNIAN
SGVLPSYFPKLTMSAEPQGLTSFALSLLLVFRTNSSYGRFDEARKIWGGILNRARNIAN
:*****:**:*****

Cre Cre06.g261750.t1.2
Vca Vocar.0002s0473.1.p

QAVTFI PAEDQAGREAVGKWTVGFTRALQAHLQEDIDLKLEKATPRWSKEEIDMLVNA
QAVTFI PAEDVAGREAVGKWA VGF CRALQAHLQEDANLREELQKQPRWSREEIDMLCSA
***** *****:*** ***** :*:**:* ***.***** *

Cre Cre06.g261750.t1.2
Vca Vocar.0002s0473.1.p

QHRPIKAISVLSSELTRQLSITQFQALQMENCTFFYDALGGCERLLRTPIPVSYRHTAR
QHRPIKAISMLSELTRQLPI SQFQALQMENVTFYDALGGCERLLRTPIPVSYRHTAR
*****:*****.*:***** *****

Cre Cre06.g261750.t1.2
Vca Vocar.0002s0473.1.p

FLTIWLAMLPLGLWERYHWSMLPVIALIGFLLLGIDEIGISIEEPFGILPLDAICGRAQT
FLTIWLSLLPLGLWDRYHWSMLPVIALIGFLLLGIDEIGISIEEPFGILPLDAICTRAQT
*****:*****:***** *****

Cre Cre06.g261750.t1.2
Vca Vocar.0002s0473.1.p

DVNSLLKEDPAVMKYVDDVRSRGRVKSPPPLPPAPAPAAA-----AAAAAARSVSP
DVVSLKDDPAVVKYISDVRQGR IAPTEPPVAGAAPVAAAPPPPPASAGGGISRSRGS
** ***:***:***:***:***:*. * * **.* ** *:*. . :* **

Cre Cre06.g261750.t1.2
Vca Vocar.0002s0473.1.p

----QPDVAKTLGLSFTNVRAGVAVAPGAPLMPQAPVRSRSPTRSVSP----SFPRASA
TAQQQPDVMKTVTSM LHNKAGIGAVAPAPRPPSPQPRARSP--RAASPGGSPFPRASA
*** ***:*: **.*:*****.* *. * *: ** *:.** *****

Cre Cre06.g261750.t1.2
Vca Vocar.0002s0473.1.p

GTGMPPVGMNGATPRVAAAPPTPPVSRPAAPAAA-----PAAGSGFTMP
GTG-----GAAA VPSPPPIKPLTSSSSSSGAVSKDSNNSTATAKPPASAPAASSA
*** ***: * .*:** * . * .:*. . * **:*... : .

Cre Cre06.g261750.t1.2
Vca Vocar.0002s0473.1.p

NFSASLSGLTGAAAAAASAA DAASSKLT KMADSMSSGAAAPAPPAAPAR-----
GFSGMFGSLADGAAAAAKSAS-AAAAKFSKIADSVVAG--TPAAPASEAKRETA AAAAMQ
.* .:***:*****: **::*:**:*: * :**.*: *

Cre Cre06.g261750.t1.2
Vca Vocar.0002s0473.1.p

-----PSTSPRPSASSPISSADADRSDSSRR---PVNWRDELQSLKATRE-----
AQPRTTPSSSSSTPSAAPANGSSDDDRSSSGRR TAAAVNWR EEL AALRAGREDAE EPASA
:* . .:*. * .*: * **.* ** .*:*** :*.* **

Cre Cre06.g261750.t1.2
Vca Vocar.0002s0473.1.p

-----PNGNGSGVAPAA--GRADADEEALRRFGNLAGRSR-----S
SASYDREFPSSWSFSSASSAAVQSGDAEDEARRRFGGLAGRARS DTTTSAAVMRGN
*... . *...** .**:* ** **.* ** .

Cre Cre06.g261750.t1.2
Vca Vocar.0002s0473.1.p

GNG-----GGSSDTELSEANRPRTRPDWRNQL
GNGLSENGYNGYGN DNGNGNTVEARGARPRTRPDWRNQL
*** * .*: * : . *****

E

RBMP2 (2/4)

Cre Cre09.g416850.t1.2 LLAAGSPPPVDA-----AAAANAAAPAAPAAAPLPADAEALGQLSEALQRELKAVVGPD
Gpe KXZ45361.1 LLSLSDGSQSSAAALVVGDAVSAASASAGGAEAAAAAAAAATQLTEALRRELAAAVGPD
Vca XP_002956061.1 -----SVATLPSASAEPAAAAAATAVSAAMAGE---GAGAD
: : : : : . * * * * : : * * * . * . *

Cre Cre09.g416850.t1.2 VDVEAAAADPSALAEAGRAVDSALGSLDSGALEALGQLPPDVRLSSLLGAVLQSALDLV
Gpe KXZ45361.1 VE----SADPAALADAAGSAITSALSSLDPAALTAVGSLPAERLRTLLEASMQSSLDLV
Vca XP_002956061.1 GQMGVYGMDF-----LDTQQLDE-----LTTALQATLQSSLDLV
: : . * * * * * * * * : : * * * : * * * : * * *

Cre Cre09.g416850.t1.2 DAAVSGVRQADSEVVGVAIVVVVLGLAIRSLVSVLGNALGGPRGGAMPASAGGGVDAA
Gpe KXZ45361.1 DAAVGGVENADSTLVGGVTAAVVALLLRSLVSMASAIISRSD-----VVAAGAGAV
Vca XP_002956061.1 EAALSGIRSADSTLVGGVTIAAVLAVVIRSLVSMAGAALSRTRG-----PGGGGGEAG
: * * : * * : * * : * * : * * : * * : * * : * * : * * : * * : * *

Cre Cre09.g416850.t1.2 GGAPRTLAEAVAAEGSARATGSRTRSALGVTALEAAAALNNEPKALLLDVRNSGDVYEQG
Gpe KXZ45361.1 GNAP---AAEAAAPGDA-----RLRPVAVGVSFAEAAATLNNDTNSLLLDIRNATDVSEQG
Vca XP_002956061.1 GGGG---GSAAAAA-----RMRP-VGVTALEAAAALNNDPQALLDIRNGSDVYEQG
* . . . *

Cre Cre09.g416850.t1.2 LADLRPFRRGSGAASAALPYLDFRTTPTLANPSGLL-----G-----
Gpe KXZ45361.1 LPDLRPFRRGAGATSVSLPYCDFRTTPTLANPSGALPSPPPAPTVAAPQAPAGLLGLLGF
Vca XP_002956061.1 LPDLRPFRRGAGATTVPLPYCDFRTTPTLANPSGSLAAAAATGPAAGSPRG-----
* . *

Cre Cre09.g416850.t1.2 ---GSGGAGAA-----GSAVAAVDPQFVPRFKQLKGLGRDSRVLLL
Gpe KXZ45361.1 GTPGSAAAGSAQPVGAAASPDGSLMAAPRGPVTVSVDPPQFQAKFKQLAPLNRDSKVFL
Vca XP_002956061.1 ---KAAAAALAPVA-----GPVTVSVDPLFCSKFKQLEGLNRDSRVFLM
: . * . : * . . : * * * * * . * * * * * * * * * * * * * * * *

Cre Cre09.g416850.t1.2 DSYGVEAPEAVALLRSDPDIERLLGGEGVSFVEGGFAGPEGWKLTGLPVMDDPEPAEAEAR
Gpe KXZ45361.1 DSYGVESPEAVSLRSDPDVERMLGSEGLAFVEGGFAGPEGWKLSGLPTAEP---ALEGE
Vca XP_002956061.1 DSYGVEAPEAVLLRSDLEVEGLLGAQGVKVFVEGGFAGPEGWK---PDLNP---LTLSS
* * * * * : * * * * * : * * * * * : * * * * * : * * * * * : * * * * * : * * * * *

Cre Cre09.g416850.t1.2 GAAGAGRPLLRRGGLDTSGLVGLLALQMRYPALLTRVLAVGAVGGVGVAAASRVWDGAV
Gpe KXZ45361.1 GAASRGLDFGR-----LAGMSGLRMRYPSLLTRVLGVGALGTGGLALASNLWDGAV
Vca XP_002956061.1 FLIFAAIQYYT-----ADISSCHVYLPFMYPSLFGRTLAVGAVGGAGVAAASSLDWASV
. : . * : * * * * : * . *

Cre Cre09.g416850.t1.2 SRGGVALAALLVADRALPTGVRPSAKLRSQQAQLEAPADSNAAAAASSQQ-PGDKRR
Gpe KXZ45361.1 ARGVGIAAAALLVADRALPTSVRPTAKLRQHLQAQLAA-----GGNGTD-----AGQRRR
Vca XP_002956061.1 SRGAVGLAAAMLLTDRVLPVPPGVRPSGKIRQYLQSQLSDPQPNDDGGSGASAAADLAAARRR
: * * . * * * * : *

Cre Cre09.g416850.t1.2 AALILRALDLVDAVGDVAVKAGQTAFAAAGGAASAA-----AKTAAASAT
Gpe KXZ45361.1 TALLLRALDLAEAVGDVAVRAGSAAAATAGSLATNAANAAYNGPYDNTATASPAANAA
Vca XP_002956061.1 TALLLRALDLAEAVGDVAVRAGGAAVAAAGSAARGAI-----AASGGGSAA
: * * . * * * * * : *

Cre Cre09.g416850.t1.2 SAAATAWPTASAGMGSEAAGDAAS-ARASTVM-----SNWRDVI DSGA-----
Gpe KXZ45361.1 AAAPAAASSAIAASAQLELTQPP-ANANTVVLPPQGMDSQQWATSA TASA-----
Vca XP_002956061.1 AAAATAANAGPADVEAELETTPPPAAAAATIVLPHGMSRQWSEAAEMAVAQVQQQPAA
: * * : * * : * * * * * : * * * * * : * * * * * : * * * * * : * * * * *

E

RBMP2 (4/4)

Cre Cre09.g416850.t1.2
Gpe KXZ45361.1
Vca XP_002956061.1

SGVEGQAQPQ-----RRAGSGRARVVVSAG-----S
-----NSNGTVRQRAALA-----S
EPSGSASESEWEAAAAATEPGDRTSLYGSVGGTASNGRTASGSIRSRNAAAVGATAAFP
* : * * . :

Cre Cre09.g416850.t1.2
Gpe KXZ45361.1
Vca XP_002956061.1

RAPSNWRQQVDG-----GSNG
RSPGNWREQVE-----GSAA
RSPSNWRVQVEGLDRPDSSTSSSSSRGFGGVPADWRTRIESGAPATATAIAADGLNGQDS
* : * . * * * * * : * . .

Cre Cre09.g416850.t1.2
Gpe KXZ45361.1
Vca XP_002956061.1

NGNGNG-----NGNGQ-----SSP--RHATPANLSPSERLAREARMRDWR
QASASV-----NGNGR-----PGTP--ARASSARLSASERAERDARLADWR
SGSAAASIGSVYDDDVSTSGNNRYSRGSYPSSPGGSRGSSRKLGAERAARTARLQDWR
..... * * * . . : * * * * : * * *

Cre Cre09.g416850.t1.2
Gpe KXZ45361.1
Vca XP_002956061.1

ARV
ARV
ARV
* * *

G

CSP41A

Cre Cre10.g440050.t1.2
Tva PNH08146.1:16-201
Gpe KUJU_scaffold_2002853
Vca Vocar.0001s1288.1.p

MQSSMRA-RVAGGARAVGTAGRRLTVKVMNSNVL IANTKGGGHAFI GLYLAKELLKKGH
-----AGRRLSVQVMNANVLIANTKGGGHAFI GLYLAKELLKKGH
MQAQLRAGRVPGGARRAVAPAGRRLTVKVMNANVLIANTKGGGHAFI GLYLAKELLKKGH
MHAQLKANR-AGGARAFAPAGRRLSVKVMNANVLIANTKGGGHAFI GLYLAKELLKKGH
*****:*:*:*:*****

Cre Cre10.g440050.t1.2
Tva PNH08146.1:16-201
Gpe KUJU_scaffold_2002853
Vca Vocar.0001s1288.1.p

KVTIMNDGSDKLTTKKNPYAKYSDLERQGLNVVWADPAKPSTYPRGTFDVVYDNNKGDLA
KVTIMNDGDESKLVKKAPFNKYSELAREGASIAWGDPAKPSTYPRGNFDVVYDNNGKDMG
KITIMNDGDESKLSKKAPFNKYSELAREGVTIAWGDPTKPKSTYPRGSFDVVYDNNGKDLA
KVTIMNDGDESKLTKKTPFSKYSELARDGATIAWGDPTKPKSTYPRGSFDVVYDNNGKDLA
*:*****. ** * *: **:* *:* .:*.**:*:*****.*****:.

Cre Cre10.g440050.t1.2
Tva PNH08146.1:16-201
Gpe KUJU_scaffold_2002853
Vca Vocar.0001s1288.1.p

SCQPLIDHFHKHKVDHYVVFSSAGAYKADPIEPMHVEGDARKSTAGHVEVEAYLEKARLPY
SCQPMIDHFHKHKVDHYVVFSSAGAYKADSIEPMHVEGDVRKATAGHVEVEAYLAQSRMPY
SCQPLIDNFKHKHVEHYVVFSSAGAYKADSIEPMHVEGDTRKSTAGHVEVEEYLKNARMPF
SCQPMIDHFHKHKVDHYVVFSSAGAYKADSIEPMHVEGDARKSTAGHVEVEAYLEKARVPY
:*.**:*:**.*****.*****.***:***** ** :*:**:

Cre Cre10.g440050.t1.2
Tva PNH08146.1:16-201
Gpe KUJU_scaffold_2002853
Vca Vocar.0001s1288.1.p

TVFQPLYIYGPNTAKDCEQWFVDRIIRD RPVLLPAPGVQLTSLTHVEDVASMLA AVPGNR
TVFQPLYIYGPNTGKDCEQWFVDRIV-----
TVFQPLYIYGPNTGKDCEQWFVDRIVRDRPVL IPTPGVQLTSVTHVEDVASMLA AVPGNR
TVFQPLYIYGPNTAKDCEQWFVDRIIRD RPVPIPSPGIQLTSLTHVEDVAAMLALVPGNR
*****.*****:

Cre Cre10.g440050.t1.2
Tva PNH08146.1:16-201
Gpe KUJU_scaffold_2002853
Vca Vocar.0001s1288.1.p

AAIGQHYNVCSDRICITFTGIAKAIGKALGKDPEIILYSPEKVGTKSGKAEGFPFRTVHF

AAIGNIYNICSDRISFVGIKSVKALGKEADVLLYSPEKVGTKSGKAEGFPFRTVHF
EAIGQMYNVCSDRICISFVGICKSVAKALGREANIVLYSPEKVGTKSGKAEGFPFRTVHF

Cre Cre10.g440050.t1.2
Tva PNH08146.1:16-201
Gpe KUJU_scaffold_2002853
Vca Vocar.0001s1288.1.p

FASADKAKRELGWKPKHDFQKDVQGLVNDYKANGRDKKEVDFSVDDKIL AALGKSVPKSS

FASSDKAKRELGWKPKHDFQKDVAAALVADYK SQGRASKEIDFSIDDKIL AALGKSVKSSA
FASSDKAKRELGWKPKHDFQKDVAAALVADYKAQGRDKKD VDFSIDDKILEALGKPFKPA

Cre Cre10.g440050.t1.2
Tva PNH08146.1:16-201
Gpe KUJU_scaffold_2002853
Vca Vocar.0001s1288.1.p

SNSSVSAS---FSRLSSSGPKAEELPRSRSSFSPPRDLKIKRTVLPANWRDSLDEDEPAK

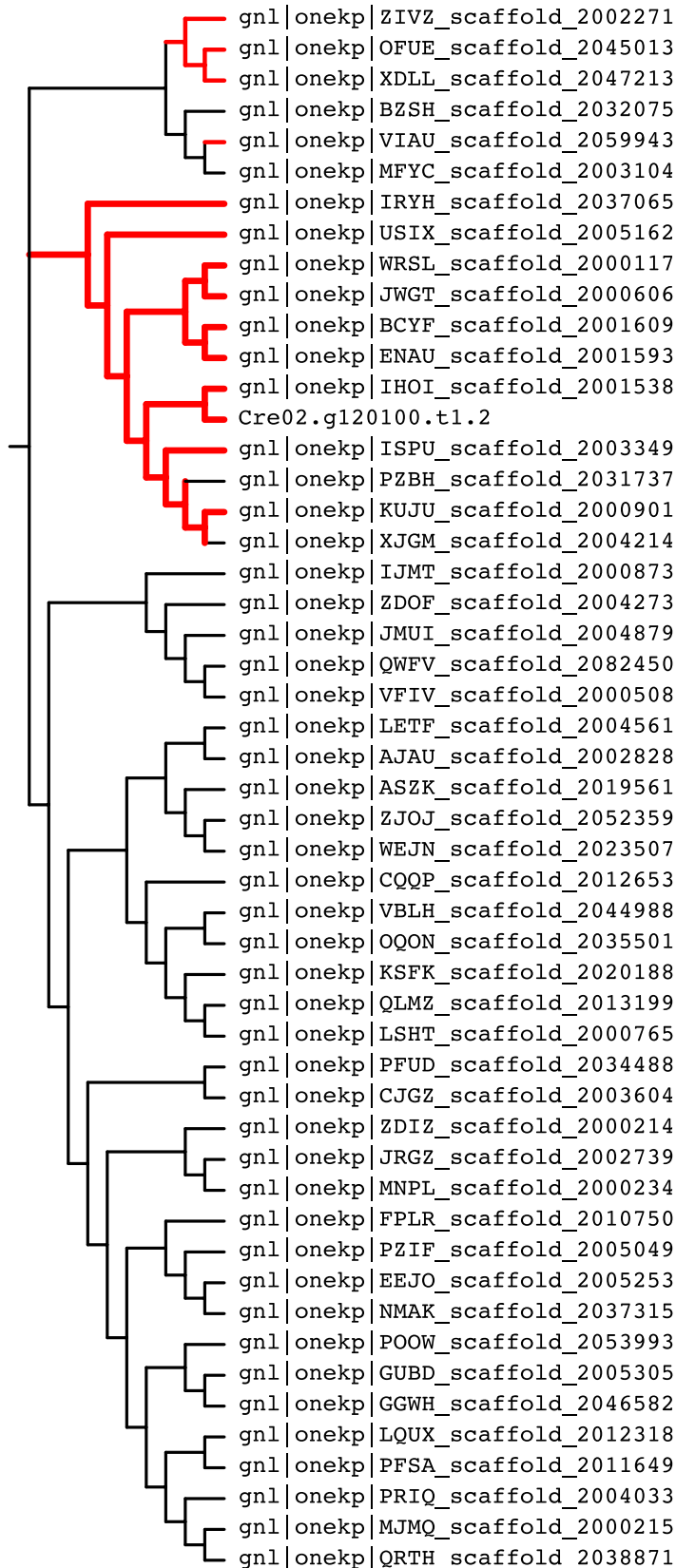
PAPSNNGAPELTRVSSPSAESGAP-RIHSSHSPRRDLKIKRTVVP SNWRDSGDGGSV--
VDSSSSSNG--FTRISSASPSPPAQRRSSFSPPRDLKIQRVLPNWRSLDGGEAVS

Cre Cre10.g440050.t1.2
Tva PNH08146.1:16-201
Gpe KUJU_scaffold_2002853
Vca Vocar.0001s1288.1.p

PAAGRSATTGRSGSVPKDWRSSL

----AASPPSRSSAPRDWRNAL
ATPSAAPRPARSSSVPRDWRSSL

H



23 24 87 90 91 94

DQQIARQVDYIVG---PSQVLKEVSQCTRA
DEQIARQVDYIVG---AMQVLKEIVACTKA
DEQIARQVDYIVG---GIQVLKEIVACTRA
DDEIARQVDYIVG---PSQVLREVNACSRG
DSQIAKQVDYIVS---ASQVLKEVSGCSRA
DSQIAKQVDYIVS---ASQVLKEVSGCSRA
DDQIAKQVDYIVG---PSQVLSEINKCSRA
DEQIAAQVDYIVA---GIQVLKEIVACTRA
DEQIAAQVDYIVA---PMQVLREVVACTRA
DEQIAAQVDYIVA---PMQVLREIVACTKA
DEQIAAQVDYIVA---PMQVLREIVACTKA
DEQIAAQVDYIVA---PMQVLREIVACTKA
DEQIAAQVDYIVA---PMQVLREIVACTKA
DEQIAAQVDYIVA---PMQVLREIVACTKA
DEQIAAQVDYIVA---PMQVLREIVACTKA
DEQIAAQVDYIVA---AAVLR EIAECKKA
DEQIAAQVDYIVA---PMQVLREIVACTKA
DEQIAAQVDYIVA---PMQVLREIVSCTKA
DGEIAQQVDYVVTG---SSQVLREIAAATRA
DDEIAEQVDYICVE---ADQVLKEVNNCTKA
DEEIGEVDYIVE---AEQVLGEIYNCTRA
DDEIAEQVDYIVD---ADQVLREVNACTKA
DDEIAEQVDYIVD---ADQVLREVNCTKA
KTQIAKQVDYIVA---PQQVLNEVDQCTRT
SSQIARQVDYIVA---PNQVLGEIDACTRA
DQQIARQVDYIVA---PNQVLGEIDACTRA
DQQIARQVDYIVA---PNQVLAEVNACKRS
DQQIARQVDYIVA---PNQVLGEIDACTRA
HGEIARQVDYIIA---PNQVLGEIDACKRA
PGEIARQVDYIIA---PHQVLAEVDACKRA
PGEIARQVDYIIA---PHQVLAEVDACKRA
DGEIARQVDYIIA---PNQVLQEVQACKRA
DGEIARQVDYIIA---PNQVLAEVNACKRA
DGEIARQVDYIIA---PNQVLAEVNACKRS
DGEIAKQVDFIVN---PSQVLREIRACVSA
NQEIARQVDYIIN---PNQVLGEVDACTRA
ADQIAKQVDYIVN---PSQVLREVSSCQIA
SDQIAAQVDYVVS---ASLVLEVDACTRA
SDQIAAQVDYVVS---ASLVLEVDACTRA
SDQIARQVDYIVG---PVQVLREIDNATKA
DDQIARQVDYIVN---PVQVLREIDNATKA
DDQIARQVDYIVN---PVQVLREIDNATKA
DDQIARQVDYIVN---PVQVLREIDNATKA
SDQIARQVDYVVG---PSAVLTEISRASKA
SDQIARQVDYVVG---PSAVLTEISRASKA
SDQIARQVDYVVG---PSAVLTEISRASKA
SDQIARQVDYIVG---PSSVLTEIARCSKA
SDQIARQVDYIVG---PSSVLTEIARCSKA
SEQIARQVDYICVG---PSAVLTEISRASKA
SEQIARQVDYICVG---PSAVLTEISRATKA
SEQIARQVDYICVG---PSAVLTEISRATKA

Fig. S9. Homologs of the Chlamydomonas Rubisco-binding pyrenoid proteins are present in the proteomes of other Volvocales. (A) For each protein, a schematic alignment of homologs is shown. Each motif is shown at its exact position in the alignment. Species abbreviation; *Cre*: *Chlamydomonas reinhardtii*; *Tso*: *Tetrabaena socialis*; *Gpe*: *Gonimum pectorale*; *Vca*: *Volvox carteri*. The current state of genome assemblies and predicted proteomes may affect the quality of the alignment. (B) Motif-containing homologs of SAGA1. (C) Motif-containing homologs of SAGA2. (D) Motif-containing homolog of RBMP1. (E) Motif-containing homologs of RBMP2. (F) Motif-containing homologs of EPYC1. (G) Motif-containing homologs of CSP41A. (H) Alignment of green algal Rubisco small subunit protein sequences. Only the amino acids of the two alpha-helices that interact with the motif are shown. The phylogenetic cladogram was constructed on full lengths sequences. Key residues identified in (21) are highlighted. Branches leading to Volvocales are colored red in the cladogram.

Table S1. Raw mass spectrometry counts of proteins coimmunoprecipitating with anti-SAGA1 antibody. (see Fig. 1E) gene_ID: as per *Chlamydomonas reinhardtii* genome v5.6 (Phytozome 13). gene_name: commonly accepted alias, when available. IPMS_WT_raw_spectral_counts. IPMS_saga1_raw_spectral_counts. pyrenoid_proteome: Y indicates presence in the pyrenoid proteome (17); N indicates absence. Rubisco_interaction_WD_score and Rubisco_interaction_Z_score are metrics of coimmunoprecipitation with RBCS-Venus-3xFLAG; WD-scores above 48 indicate particularly high confidence (see 18).

Table S2. *Chlamydomonas reinhardtii* proteins containing one or more instances of the motif. gene_ID: as per *Chlamydomonas reinhardtii* genome v5.6 (Phytozome 13). gene_name: commonly accepted alias, when available. #motifs: number of motifs. motif_scores: from 0 to 6, as defined in Material and Methods. motif_positions: position is for the 9th amino acid upstream of the motif's tryptophan. motif_sequences: spanning the colored residues in Fig. S2A. motif_disorder_scores: disorder scores used in the analysis shown in Fig. S7 (motif disorder scores were only determined for motifs whose motif_score was equal to or greater than 3). pyrenoid_proteome: Y indicates presence in the pyrenoid proteome (17); N indicated absence. chloroplast_proteome: Y indicates presence in the chloroplast proteome (50); N indicates absence. defline/description/synonyms: succinct gene descriptions as per Phytozome. predalgo_protein_localization: predicted subcellular localization, as per (52), where C = chloroplast, M = mitochondria, N = nucleus, SP = secretory pathway, O = other. PFAM/Panther/KOK/KEGG/Gene_Ontology_term: different predicted functional annotations and gene ontologies. best_arabidopsis_TAIR_hit_name/symbol/defline: available data for the closest homolog in the model land plant *Arabidopsis thaliana*.

REFERENCES AND NOTES

1. J. A. Raven, Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments. *Aquat. Microb. Ecol.* **56**, 177–192 (2009).
2. A. Bar-Even, E. Noor, R. Milo, A survey of carbon fixation pathways through a quantitative lens. *J. Exp. Bot.* **63**, 2325–2342 (2012).
3. C. J. Still, J. A. Berry, G. J. Collatz, R. S. DeFries, Global distribution of C₃ and C₄ vegetation: Carbon cycle implications. *Global Biogeochem. Cycles* **17**, 6-1–6-14 (2003).
4. E. Granum, J. A. Raven, R. C. Leegood, How do marine diatoms fix 10 billion tonnes of inorganic carbon per year? *Can. J. Bot.* **83**, 898–908 (2005).
5. J. A. Raven, J. Beardall, M. Giordano, Energy costs of carbon dioxide concentrating mechanisms in aquatic organisms. *Photosynth. Res.* **121**, 111–124 (2014).
6. S. C. Maberly, B. Gontero, Ecological imperatives for aquatic CO₂-concentrating mechanisms. *J. Exp. Bot.* **68**, 3797–3814 (2017).
7. C. B. Field, M. J. Behrenfeld, J. T. Randerson, P. Falkowski, Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).
8. J. H. Hennacy, M. C. Jonikas, Prospects for engineering biophysical CO₂ concentrating mechanisms into land plants to enhance yields. *Annu. Rev. Plant Biol.* **71**, 461–485 (2020).
9. B. D. Engel, M. Schaffer, L. Kuhn Cuellar, E. Villa, J. M. Plitzko, W. Baumeister, Native architecture of the *Chlamydomonas* chloroplast revealed by in situ cryo–electron tomography. *eLife* **4**, e04889 (2015).
10. M. T. Meyer, C. Whittaker, H. Griffiths, The algal pyrenoid: Key unanswered questions. *J. Exp. Bot.* **68**, 3739–3749 (2017).
11. J. D. Pickett-Heaps, *Green Algae: Structure, Reproduction and Evolution in Selected Genera* (Sinauer Associates, 1975).

12. C. Toyokawa, T. Yamano, H. Fukuzawa, Pyrenoid starch sheath is required for LCIB localization and the CO₂-concentrating mechanism in green algae. *Plant Physiol.* **182**, 1883–1893 (2020).
13. L. C. M. Mackinder, M. T. Meyer, T. Mettler-Altmann, V. K. Chen, M. C. Mitchell, O. D. Caspari, E. S. Freeman Rosenzweig, L. Pallesen, G. Reeves, A. K. Itakura, R. Roth, F. Sommer, S. Geimer, T. Mühlhaus, M. Schroda, U. W. Goodenough, M. Stitt, H. Griffiths, M. C. Jonikas, A repeat protein links rubisco to form the eukaryotic carbon-concentrating organelle. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5958–5963 (2016).
14. A. K. Itakura, K. X. Chan, N. Atkinson, L. Pallesen, L. Wang, G. Reeves, W. Patena, O. D. Caspari, R. Roth, U. W. Goodenough, A. J. McCormick, H. Griffiths, M. C. Jonikas, A Rubisco-binding protein is required for normal pyrenoid number and starch sheath morphology in *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 18445–18454 (2019).
15. E. S. Freeman Rosenzweig, B. Xu, L. Kuhn Cuellar, A. Martinez-Sanchez, M. Schaffer, M. Strauss, H. N. Cartwright, P. Ronceray, J. M. Plitzko, F. Förster, N. S. Wingreen, B. D. Engel, L. C. M. Mackinder, M. C. Jonikas, The eukaryotic CO₂-concentrating organelle is liquid-like and exhibits dynamic reorganization. *Cell* **171**, 148–162.e19 (2017).
16. T. Wunder, S. L. H. Cheng, S.-K. Lai, H.-Y. Li, O. Mueller-Cajar, The phase separation underlying the pyrenoid-based microalgal Rubisco supercharger. *Nat. Commun.* **9**, 5076 (2018).
17. Y. Zhan, C. H. Marchand, A. Maes, A. Mauries, Y. Sun, J. S. Dhaliwal, J. Uniacke, S. Arragain, H. Jiang, N. D. Gold, V. J. J. Martin, S. D. Lemaire, W. Zerges, Pyrenoid functions revealed by proteomics in *Chlamydomonas reinhardtii*. *PLOS ONE* **13**, e0185039 (2018).
18. L. C. M. Mackinder, C. Chen, R. D. Leib, W. Patena, S. R. Blum, M. Rodman, S. Ramundo, C. M. Adams, M. C. Jonikas, A spatial interactome reveals the protein organization of the algal CO₂-concentrating mechanism. *Cell* **171**, 133–147.e14 (2017).
19. M. E. Baker, W. N. Grundy, C. P. Elkan, Spinach CSP41, an mRNA-binding protein and ribonuclease, is homologous to nucleotide-sugar epimerases and hydroxysteroid dehydrogenases. *Biochem. Biophys. Res. Commun.* **248**, 250–254 (1998).

20. N. Atkinson, C. N. Velanis, T. Wunder, D. J. Clarke, O. Mueller-Cajar, A. J. McCormick, The pyrenoidal linker protein EPYC1 phase separates with hybrid *Arabidopsis*–*Chlamydomonas* Rubisco through interactions with the algal rubisco small subunit. *J. Exp. Bot.* **70**, 5271–5285 (2019).
21. S. He, H. T. Chou, D. Matthies, T. Wunder, M. T. Meyer, N. Atkinson, A. Martinez-Sanchez, P. D. Jeffrey, S. A. Port, W. Patena, G. He, V. K. Chen, F. M. Hughson, A. J. McCormick, O. Mueller-Cajar, B. D. Engel, Z. Yu, M. C. Jonikas, The structural basis of rubisco phase separation in the pyrenoid. bioRxiv 2020.08.16.252809 [Preprint]. 16 August 2020.
<https://doi.org/10.1101/2020.08.16.252809>.
22. S. F. Banani, A. M. Rice, W. B. Peeples, Y. Lin, S. Jain, R. Parker, M. K. Rosen, Compositional control of phase-separated cellular bodies. *Cell* **166**, 651–663 (2016).
23. J.-M. Choi, A. S. Holehouse, R. V. Pappu, Physical principles underlying the complex biology of intracellular phase transitions. *Annu. Rev. Biophys.* **49**, 107–133 (2020).
24. E. W. Martin, A. S. Holehouse, I. Peran, M. Farag, J. J. Incicco, A. Bremer, C. R. Grace, A. Soranno, R. V. Pappu, T. Mittag, Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694–699 (2020).
25. N. A. Pronina, V. E. Semenenko, Localization of membrane bound and soluble carbonic anhydrase in the *Chlorella* cell. *Fiziol. Rastenii* **31**, 241–251 (1984).
26. J. A. Raven, CO₂-concentrating mechanisms: A direct role for thylakoid lumen acidification? *Plant Cell Environ.* **20**, 147–154 (1997).
27. U. W. Goodenough, R. P. Levine, Chloroplast structure and function in *ac-20*, a mutant strain of *Chlamydomonas reinhardtii*. 3. Chloroplast ribosomes and membrane organization. *J. Cell Biol.* **44**, 547–562 (1970).
28. O. D. Caspari, M. T. Meyer, D. Tolleter, T. M. Wittkopp, N. J. Cunniffe, T. Lawson, A. R. Grossman, H. Griffiths, Pyrenoid loss in *Chlamydomonas reinhardtii* causes limitations in CO₂ supply, but not thylakoid operating efficiency. *J. Exp. Bot.* **68**, 3903–3913 (2017).

29. A. Mukherjee, C. S. Lau, C. E. Walker, A. K. Rai, C. I. Prejean, G. Yates, T. Emrich-Mills, S. G. Lemoine, D. J. Vinyard, L. C. M. Mackinder, J. V. Moroney, Thylakoid localized bestrophin-like proteins are essential for the CO₂ concentrating mechanism of *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 16915–16920 (2019).
30. J. C. Villarreal, S. S. Renner, Hornwort pyrenoids, carbon-concentrating structures, evolved and were lost at least five times during the last 100 million years. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18873–18878 (2012).
31. J. A. Raven, J. Beardall, P. Sánchez-Baracaldo, The possible evolution and future of CO₂-concentrating mechanisms. *J. Exp. Bot.* **68**, 3701–3716 (2017).
32. H. Griffiths, M. T. Meyer, R. E. M. Rickaby, Overcoming adversity through diversity: Aquatic carbon concentrating mechanisms. *J. Exp. Bot.* **68**, 3689–3695 (2017).
33. M. D. Herron, J. D. Hackett, F. O. Aylward, R. E. Michod, Triassic origin and early radiation of multicellular volvocine algae. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3254–3258 (2009).
34. M. T. Meyer, T. Genkov, J. N. Skepper, J. Jouhet, M. C. Mitchell, R. J. Spreitzer, H. Griffiths, Rubisco small-subunit α -helices control pyrenoid formation in *Chlamydomonas*. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 19474–19479 (2012).
35. B. D. Rae, B. M. Long, B. Förster, N. D. Nguyen, C. N. Velanis, N. Atkinson, W. Y. Hee, B. Mukherjee, G. D. Price, A. J. McCormick, Progress and challenges of engineering a biophysical CO₂-concentrating mechanism into higher plants. *J. Exp. Bot.* **68**, 3717–3737 (2017).
36. L. C. M. Mackinder, The *Chlamydomonas* CO₂-concentrating mechanism and its potential for engineering photosynthesis in plants. *New Phytol.* **217**, 54–61 (2018).
37. R. Zhang, W. Patena, U. Armbruster, S. S. Gang, S. R. Blum, M. C. Jonikas, High-throughput genotyping of green algal mutants reveals random distribution of mutagenic insertion sites and endonucleolytic cleavage of transforming DNA. *Plant Cell* **26**, 1398–1409 (2014).

38. J. Kropat, A. Hong-Hermesdorf, D. Casero, P. Ent, M. Castruita, M. Pellegrini, S. S. Merchant, D. Malasarn, A revised mineral nutrient supplement increases biomass and growth rate in *Chlamydomonas Reinhardtii*. *Plant J.* **66**, 770–780 (2011).
39. M. C. Mitchell, M. T. Meyer, H. Griffiths, Dynamics of carbon-concentrating mechanism induction and protein relocalization during the dark-to-light transition in synchronized *Chlamydomonas reinhardtii*. *Plant Physiol.* **166**, 1073–1082 (2014).
40. D. Strenkert, S. Schmollinger, S. D. Gallaher, P. A. Salomé, S. O. Purvine, C. D. Nicora, T. Mettler-Altmann, E. Soubeyrand, A. P. M. Weber, M. S. Lipton, G. J. Basset, S. S. Merchant, Multiomics resolution of molecular events during a day in the life of *Chlamydomonas*. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 2374–2383 (2019).
41. R. J. Porra, W. A. Thompson, P. E. Kriedemann, Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: Verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochim. Biophys. Acta Bioenerg.* **975**, 384–394 (1989).
42. K. Hilpert, D. F. H. Winkler, R. E. W. Hancock, Peptide arrays on cellulose support: SPOT synthesis, a time and cost efficient method for synthesis of large numbers of peptides in a parallel and addressable fashion. *Nat. Protoc.* **2**, 1333–1349 (2007).
43. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
44. One Thousand Plant Transcriptomes Initiative, One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).
45. F. Madeira, Y. M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. N. Tivey, S. C. Potter, R. D. Finn, R. Lopez, The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **47**, W636–W641 (2019).

46. E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, A. Bairoch, ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–3788 (2003).
47. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
48. K. Hofmann, W. Stoffel, TMBase—A database of membrane spanning proteins segments. *Biol. Chem.* **374**, 166 (1993).
49. T. Yang, Q. Liu, B. Kloss, R. Bruni, R. C. Kalathur, Y. Guo, E. Kloppmann, B. Rost, H. M. Colecraft, W. A. Hendrickson, Structure and selectivity in bestrophin ion channels. *Science* **346**, 355–359 (2014).
50. M. Terashima, M. Specht, M. Hippler, The chloroplast proteome: A survey from the *Chlamydomonas reinhardtii* perspective with a focus on distinctive features. *Curr. Genet.* **57**, 151–168 (2011).
51. B. Mészáros, G. Erdős, Z. Dosztányi, IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337 (2018).
52. M. Tardif, A. Atteia, M. Specht, G. Cogne, N. Rolland, S. Brugiè, M. Hippler, M. Ferro, C. Bruley, G. Peltier, O. Vallon, L. Cournac, PredAlgo: A new subcellular localization prediction tool dedicated to green algae. *Mol. Biol. Evol.* **29**, 3625–3639 (2012).