

**Evolutionary Maintenance of the PTS2 Protein Import Pathway
in the Stramenopile Alga *Nannochloropsis***

Journal: *Frontiers in Cell and Developmental Biology*;

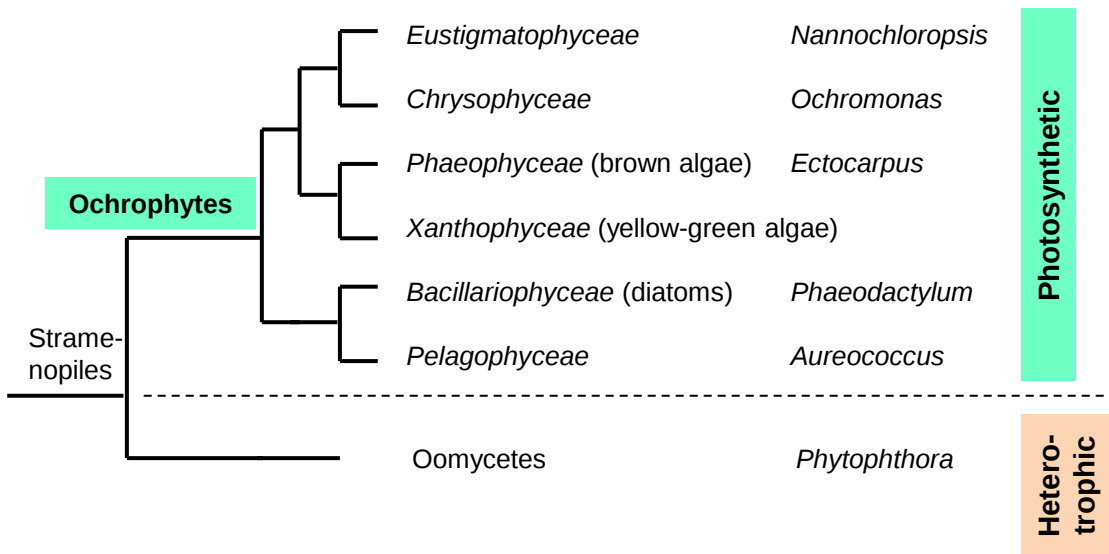
Authors: Dmitry Kechasov, Imke de Grahl, Pierre Endries and Sigrun Reumann;

Affiliation: Plant Biochemistry and Infection Biology, Institute of Plant Science and Microbiology, Universität Hamburg, D-22609 Hamburg, Germany;

Email address of corresponding author: sigrun.reumann@uni-hamburg.de

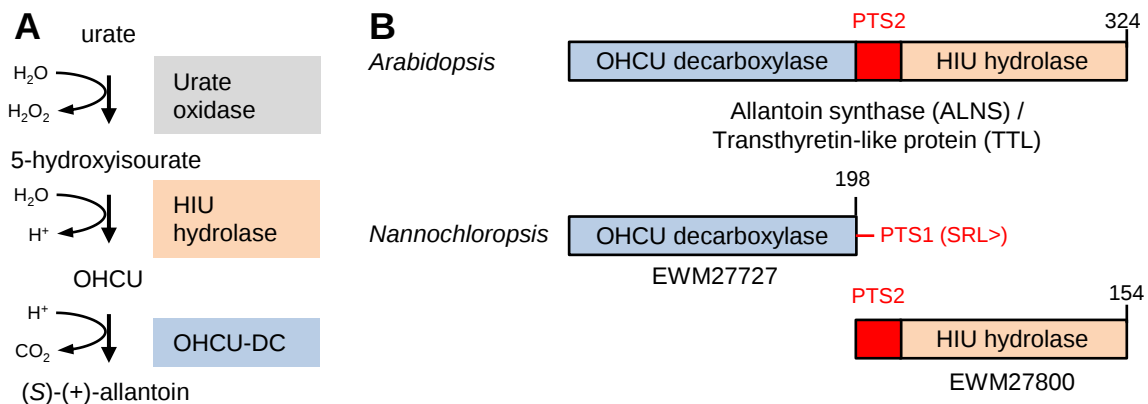
Supplementary Figures

Suppl. Figure S1



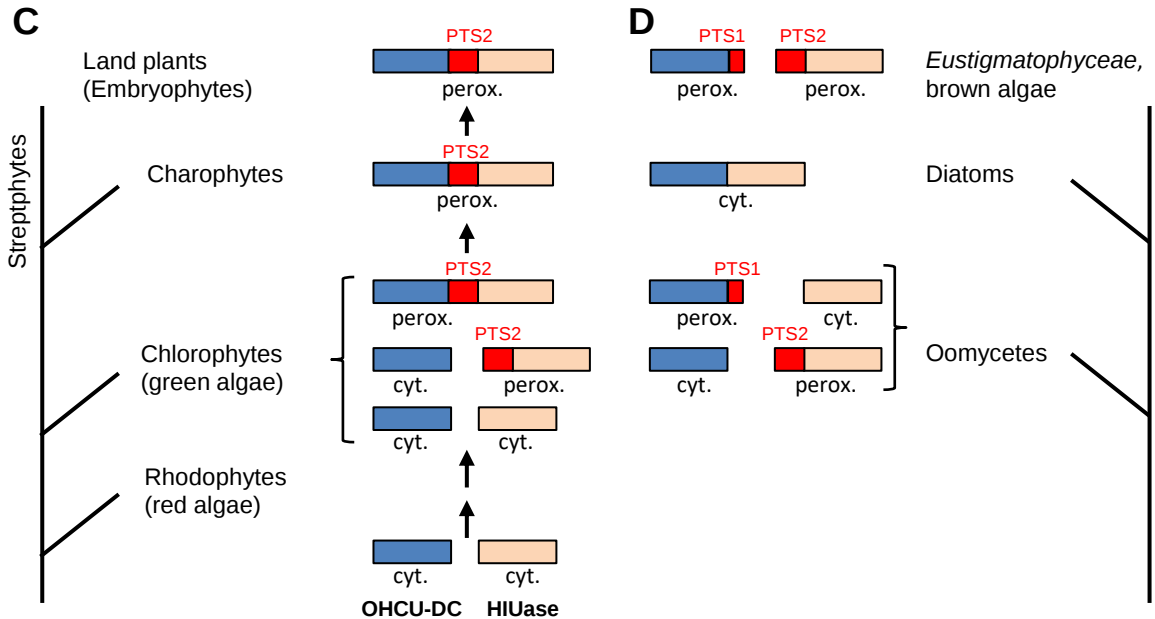
Suppl. Figure S1: Schematic phylogenetic tree of stramenopiles according to multiple gene analyses by Yoon et al. (2012). Stramenopiles contain both photosynthetic members (the ochrophytes), which possess complex plastids of red algal origin, and aplastidic and non-photosynthetic members (e.g. oomycetes). Ochrophytes include many ecologically important lineages (diatoms, kelps, pelagophytes) and *Nannochloropsis* as a potential model lineages for biofuels research. Ochrophytes form the most significant component of eukaryotic marine phytoplankton (Dorrel et al., 2017).

Suppl. Figure S2 A, B



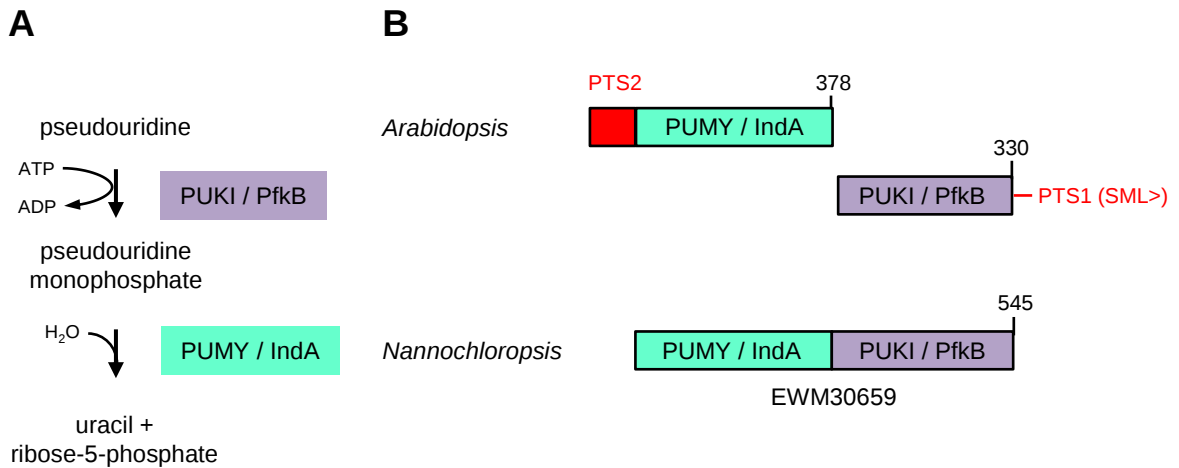
Suppl. Figure S2: Activity, domain architecture and evolution of peroxisomal enzymes of urate catabolism in *A. thaliana* and *N. gaditana*. **(A)** Urate degradation to (S)-(+)-allantoin is catalyzed by urate oxidase, 5-hydroxyisourate (HIU) hydrolase and (S)-2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline (OHCU) decarboxylase (DC). **(B)** The domain architecture of OHCU-DC and HIU hydrolase shows for *A. thaliana* a bifunctional fusion protein with an unusual internal PTS2 and for *N. gaditana* two distinct PTS1- or PTS2-carrying enzymes. **(C)** and **(D)** are provided on the next page.

Suppl. Figure S2 C, D



Suppl. Figure S2: Activity, domain architecture and evolution of peroxisomal enzymes of urate catabolism in *A. thaliana* and *N. gaditana*. (A) and (B) are shown on the previous page. (C) Schematic presentation of enzyme evolution in *Viridiplantae*. Initially both enzymes (OHCU-DC and HIUase) were separated and cytosolic, as still nowadays in Rhodophytes and few species of chlorophytes (*Trebouxiophyceae*). For instance, *Chlorella sorokiniana* still uses the ancient forms of single cytosolic enzymes. In the last common ancestor of Chlorophyceae and Streptophytes, HIUase first evolved a PTS2 (as found today in some *Trebouxiophyceae*). Examples include HIUase of *Micractinium conductrix* (RAX₅QL) and *Chlorella variabilis* (RAX₅QL). Subsequently, HIUase merged N-terminally with OHCU-DC, thereby leading to the unusual internal PTS2. Gene fusions (OHCU-DC-HIUase), all with the internal PTS2, are indeed detectable in two *Trebouxiophyceae* species (*Trebouxia* sp. A1-2, *Coccomyxa subellipsoidea*, both RAX₅HL). (D) Schematic presentation of enzyme evolution in Stramenopiles with different variants in *Nannochloropsis* (*Eustigmatophyceae*, OHCU-DC with PTS1, HIUase with PTS2), diatoms (cytosolic protein fusion) and Oomycetes (separate enzymes with generally only one PTS-carrying enzyme). The constellation in Oomycetes implies peroxisome import of the other enzyme by oligomerization and piggy-back mechanism.

Suppl. Figure S3



Suppl. Figure S3: Comparison of pseudouridine degradation catalyzed by single peroxisomal (*A. thaliana*) or a bifunctional cytosolic fusion enzyme (*N. gaditana*). **(A)** In *Arabidopsis*, pseudouridine degradation to uracil was recently demonstrated to be catalyzed by two peroxisomal enzymes, pseudouridine kinase (PUKI/PfkB) and pseudouridine monophosphate glycosylase (PUMY/IndA, Chen and Witte, 2020). **(B)** Comparative schematic diagram of the protein structures of PUKI/PfkB and PUMY/IndA in *A. thaliana* (two distinct PTS1- or PTS2-carrying enzymes) and *N. gaditana* (bifunctional fusion protein) without predicted PTS1/2 (but the decapeptide RLx₆HV).

Suppl. Figure S4

A: Thiolase (NgPKT)

N_gad_EWM24705	1	-----MSSTQTSQANKRIEHLISGHLVKKTSRPG
N_sal_TFJ86767	1	-----MSSTQTSQANKRIEHLISGHLVKKTSRPG
A_eut_KAF073646	1	-----MQ---RIERIQQHLQ--APSSA
A_hyp_OQR92766	1	-----MQ---RIERIQGHMR---PAAQ
G_spl_KAF132289	1	-----MD---RIERIRSHVQQNKPSAA
A_lai_CCA20150	1	-----MERVRHLRSHLT---AKA
P_inf_KAF404366	1	-----MD---RINRIRSHVANGSPAPA
A_ste_KAF069031	1	-----MH---RIERIQQHLQ--APGSA
T_cla_QR95928	421	SLLHEPVPDVDNIPTMQRIERIQGHMH--APAA-
E_sp_CAB111896	1	-----MEAARMEVIGSHLS--ASSGA
H_fer_GBG28030	1	-----MANRLEKLA AHLQ--VAGSA
C_roe_KAA016020	6	VVLTEGAEALRAHAAASRVEALARHLGSAAASGP

B: HIT1 (NgHIT1)

N_gad_EWM29206	1	MSQRLVRLSQHLLVH-
A_lai_CCA14803	1	--MRLIRFAEHATSS-
P_oli_TMW57389	1	MLHRLRVLNSHLAASS

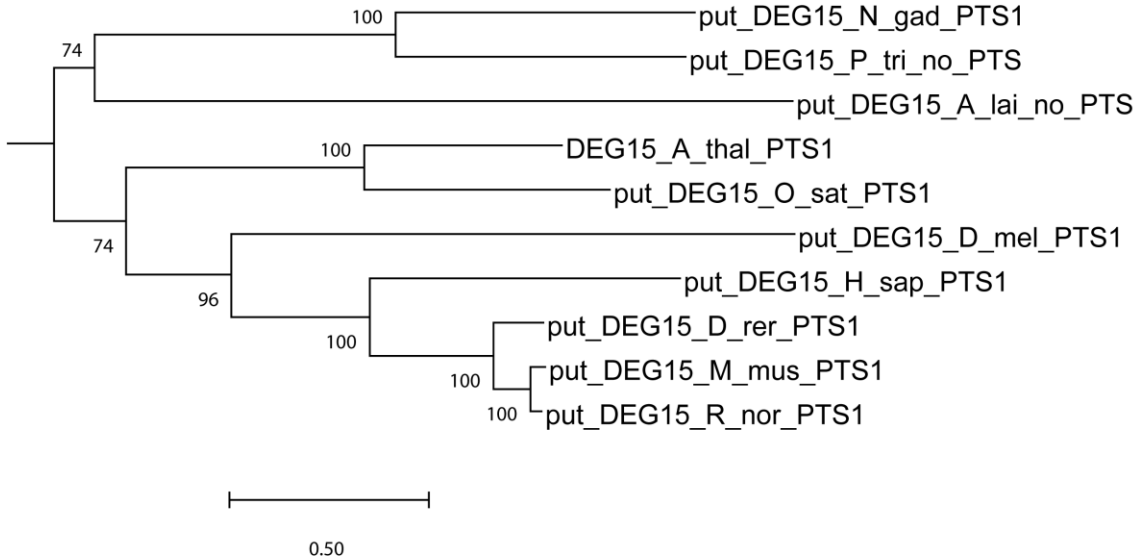
C: HIUase (NgHIUase)

N_gad_EWM27800	1	-----MPAS-STTPATSRLTALAAHLSSV
P_bra_TYZ59927	1	-----M-----RGSADRLASTIQSHLQA-
E_sp_CAB110310	1	-----MPNTPQERLDRITNHLAAG
G_spl_KAF133578	1	-----MPTSAPSPSSTRIRISTLQQHLQM-
S_par_XP_012204	1	-----M-----N-----RLEVVAKHVMP-
S_dic_XP_008610	1	-----M-----N-----RLEVIAQHVMP-
P_cac_KAF179436	841	CSTGFNTWDLAVLVPNLHLSVTLMSSPSRRVNAVNRHLSS-
P_hal_XP_024577	1	-----MASRINAVQRHASA-
A_ste_KAF071910	1	-----M-----ERLQTIQSHMAA-
A_eut_KAF074330	1	-----M-----ERLRRIQNHVAM-

Suppl. Figure S4: Analysis of PTS2 conservation of predicted *N. gaditana* PTS2 proteins in Stramenopiles. Putative orthologs of the PTS2-carrying thiolase (NgPKT, A), NgHIT1 (B) and HIUase (C) were identified by protein BLAST searches at NCBI using Genbank focusing on Stramenopiles. Homologs were aligned by COBALT (Papadopoulos and Agarwala, 2007) and sequence conservation labeled by Boxshade. Acronyms: A_eut, *Aphanomyces euteiches*; A_hyp, *Achlya hypogyna*; A_lai, *Albugo laibachii* Nc14; A_ste, *Aphanomyces stellatus*; C_roe, *Cafeteria roenbergensis*; E_sp, *Ectocarpus* sp. CCAP 1310/34; G_spl, *Globisporangium splendens*; H_fer, *Hondaea fermentalgiana*; N_gad, *Nannochloropsis gaditana* B-31; N_sal, *Nannochloropsis salina* CCMP1776; P_bra, *Pythium brassicum*; P_cac, *Phytophthora cactorum*; P_hal, *Plasmopara halstedii*; P_inf, *Phytophthora infestans*; P_oli, *Pythium oligandrum*; S_dic, *Saprolegnia diclina* VS20; S_par, *Saprolegnia parasitica* CBS 223.65; T_cla, *Thraustotheca clavata*.

Suppl. Figure S5 A

A



Suppl. Figure S5: Phylogenetic and homology analysis of DEG15 in *A. thaliana*, Stramenopiles and other eukaryotes. (A) Homologs of the PTS1-carrying DEG15 homolog of *N. gaditana* (putNgDEG15, EWM21659, SLL>) were identified by protein BLAST searches at NCBI using Genbank and focusing on Stramenopiles, green and red algae, land plants and animals. Sequences were aligned by MUSCLE (Edgar 2004) and the phylogenetic tree was constructed using the Bayesian inference method (Ronquist 2012). The branch support values were calculated as Bayesian posterior probabilities and are shown next to the branches. The platform Phylogeny.fr (Dereeper et al., 2008) was used for phylogenetic analysis, and MEGA X was used for tree visualization (Kumar, 2018). **(B)** is shown on the next page.

Suppl. Figure S5 B

B

```

put_DEG15_A_lai 327 IIAKKINV--VRIQGTSS--TIERLPAVEIGNSNNVEFGDEWWMFGNPSS--GH--NTITV
put_DEG15_D_mel 351 DSDKPEIDALITAPQDVPQ---CCVRLASPA-TVGQMVMYNAGHPY--V--FSFRHDF
put_DEG15_P_tri 211 DVHPALDVAIVKTI-VTPKTSPPWLA--SFCASSD-----LVGQGI--IAIG----PFGI
put_DEG15_N_gad 317 GFDPDKDVAVIRV--EAPPAS---LRPLFVCGSSST-----LKVGQALAI--PFGI
DEG15_A_thal 485 ICKEQLDIALIQL--PY--PGK---LQPTANFSSP-PLCTTAHVVGHGFL--GP--RCGL
put_DEG15_O_sat 502 ISKGFIDVAILIQ--EKTPIE---LCA--RPEFVCP-TAGSSVYVVGHGFL--GP--RSGL
put_DEG15_D_rer 362 VVSSPDIADVLELQALTK---RTERETK--FHT---GEDVVVVGNGAL--GS--RCG
put_DEG15_H_sap 402 QETCPDIADVSLLEDLDV---PIPVPAE--FHE---GEAVSVVVGNGV--G--SCG
put_DEG15_M_mus 404 QETSPDIADVSLDEELNGV---PTVPVPAE--FHE---GEPVSVVVGNGV--G--ACG
put_DEG15_R_nor 403 QETSPDIADVSLDEELNGV---PVVPVPAE--FHE---GEPVSVVVGNGV--G--ACG

put_DEG15_A_lai 381 HHAICSGTDSQVYDQGQIDKA-----MLRTAAQ--DNFGSGGAA--DK-K
put_DEG15_D_mel 404 NPSTFCG---RVIKCDTG-----A--MSDGSVQAGSGSGGFMFDQ-N
put_DEG15_P_tri 260 DNTVTVGVVSAINREL--RGND-----GMVSPF--RNC--QTDCA--NPGNSGGPL--NL-K
put_DEG15_N_gad 363 DHLLTSGVVISGLGREVR--PSGR-----P--SNV--QTDAA--NPGNSGGPL--DS-A
DEG15_A_thal 535 SPSTCSGVAKVVF--AKRRRLNTQ---SISQEVAEFPA--MLETTAAVHPGSGGGA--NS-S
put_DEG15_O_sat 552 SSLSLSSGVSKIV--IPSP--QHSQLSSVVEVNNMDIPV--MLQTTAAVHPGASGGV--LDS-L
put_DEG15_D_rer 410 -PSLTSGILSRVITHQSQ-----PV--MLQTTCAVQSGASGGAV--RS-DT
put_DEG15_H_sap 450 -PSVTSGLISAVVQVNGT-----PV--MLQTTCAVHSGSGGGLF--FS-NHS
put_DEG15_M_mus 452 -PSVTSGLISAVVRVDGS-----PV--MLQTTCAVHGGSGSGGLF--FS-SGS
put_DEG15_R_nor 451 -PSVTSGLISAVVHVDDA-----PV--MLQTTCAVHGGSGSGGLF--FS-RS

put_DEG15_A_lai 423 GGLNG-----SFS-----LRODR--R-----A--NMV
put_DEG15_D_mel 440 GCLLGCVSNIKLDD--VVYPN--NTATPLCD--RNTLQQ--ARTND--N-----VLSNL
put_DEG15_P_tri 312 GEVVG---NTAIVTSSGSSAG--GFA--PSDEKDV--ERMIRTD--KKGTQYQAWLGLAIV
put_DEG15_N_gad 410 GRLVG---NTAIYSPS--ASAG--GFAIPVDTLKFVETIIKDGK--VR-----PLIG--TYL
DEG15_A_thal 589 GHMIGLVTSNAR--HGAGTVI--PHLNF--SIPCAVLA--PIF--FAEDMONTT-----ILQTL
put_DEG15_O_sat 609 GRMVGILTSNAR--HGGGSI--PHLNF--SIPCKSLEMV--F--YSAGKDFK-----ILEQL
put_DEG15_H_sap 451 GDLGLGVSSNTRDYAAKVTPH--LNF--SIPVTL--EPIL--RFAQTGDAA-----VFNVL
put_DEG15_M_mus 491 GNLLGILTSNTRDNNTGT--YPH--LNF--SIPVTL--QPALQ--QYSQTGD--G-----GLREL
put_DEG15_R_nor 492 GDLGLIVASNTRDNNTGT--YPH--LNF--SIPVTL--QPALQ--QYSQTGD--C-----GLREL

put_DEG15_A_lai 446 KGAIEVA--KRMHML-----
put_DEG15_D_mel 489 -VASPDV--H--VWSLEMP--
put_DEG15_P_tri 369 KATSNCTLGSN--NVAKVMRKSPAAEAGVQAIRVFEQDASVQYGDAVDLDRVPGEKVALT
put_DEG15_N_gad 462 SSQAKALGR--GGR--GRA-----GGRAGSAR
DEG15_A_thal 639 DQPSFEI--SS--WALMP--SLSPKTEQSLPNLPKLLKDGNNKQTKGQFAKFIAETQDMFVK
put_DEG15_O_sat 658 DKPNEV--SSVWAL--PTS-----SPFFSTSPENGRGKGKVLFSKFLADKQEGLS
put_DEG15_D_rer 502 DSAVEDV--R--VVR--LQN-----
put_DEG15_H_sap 542 DRAAEPV--R--VVR--LQRPL-----
put_DEG15_M_mus 544 DHTTEPV--R--VVR--LQRPL-----
put_DEG15_R_nor 543 DHTTEPV--R--VVR--LQRPL-----

put_DEG15_A_lai -----
put_DEG15_D_mel 505 -----RSKL-----
put_DEG15_P_tri 429 LEDSSGD--RRV--YITLCTRP
put_DEG15_N_gad 487 ARALT---SRR--FSKL-----
DEG15_A_thal 697 PTKLS---R--V--PSKL-----
put_DEG15_O_sat 707 IKDIEAFLR--R--PSKL-----
put_DEG15_D_rer 517 -----QSKL-----
put_DEG15_H_sap 559 -----A--A--RSKL-----
put_DEG15_M_mus 561 -----S--V--RSKL-----
put_DEG15_R_nor 560 -----P--A--RSKL-----

```

Suppl. Figure S5: Phylogenetic and homology analysis of DEG15 in *A. thaliana*, stramenopiles and other eukaryotes. (A) is shown on the previous page. (B) The alignment was performed with MUSCLE (v3.8.31, Edgar 2004). Only the conserved C-terminal domain is shown (amino acid residues 485 to 720 of AtDEG15). Acronyms: A_lai, *Albugo laibachii* Nc14 (CCA14173.1); A_thal, *Arabidopsis thaliana* (NP_174153.2); D_mel, *Drosophila melanogaster* (NP_611968.1); D_rer, *Danio rerio* (NP_001122182.1); H_sap, *Homo sapiens* (NP_775826.2); P_tri, *Phaeodactylum tricornutum* CCAP 1055/1 (XP_002183960.1); M_mus, *Mus musculus* (NP_082188.1); N_gad, *Nannochloropsis gaditana* B-31 (EWM21659.1); O_sat, *Oryza sativa* Japonica (XP_015640602.1); R_nor, *Rattus norvegicus* (NP_001102402.1).

Suppl. Figure S6

A. PEX7 (NgPEX7, EWM28214.1) :

MKRESFHAANTGGGGLGGPIPDVAFQTQFQCCS (intron 1) VEFSPFQEARLAVATAQYFGIIG
NGRQHILE (intron 2) IGPDGNLREIRSFLTQEGLYDCCWSEANQNQLVSASADGSLKLWDVMT
SDGYPAVH (intron 3) WQEHSAEVSSVHWNQVVKTNFLSASWDGSIKLWDPHHPTSLSTYCGHT
GCVYAGIHSRPHRFLSCGTGDSLRIWDTKLPPSHATSLGLGRAEGGAVQVVRAHEGEVLSADWDKY
QDFLVYTGGVDRSIKIWDLRRPSLPLGFLHGHGYAVRRLKTSPHQEGVLGSVSYDMSCRVWGPTRAGG
RAGELWRCEEHTEFVQGLDFHLFWPGRIATCGWDRRCVVTLP

B. Malate synthase 2 (NgMLS2, EWM30341.1) :

MFLSGSMAQ**RIQVISRHL**QKKAFEEKNGATRRLTSFEAGPHI (intron 1) IEYDADVDGLHSS
IVDEVLTGALSFLAELVQHFNKDVLE (intron 2) LYRRRAEVQTKIETGTYEFGFSPETAQIRK
(intron 3) GIWQVEPVPSVLLDRRVDVGDVAPDDARALVSALNSGAQGVQCDFFDGCPTWRNV
LLGIRNVMDASQGVLCYPDSVRAHSVAIVSNAAVMMLRPRAWCMTEMHFYVNGRAIPGPLLDY (int-
ron 4) GLLIYHCGAELEMRGKGPFFYCSKVENYLEARLWNSIFTWSEVKLGLKK (intron 5)
NSVKACVLIENITASFQMDIILYELRNHSAGLNCGIWDYSASFIAQFARSDMIFPDRTKYVTMECGF
MRNYMRLLVDTCHKRGAIATTGMAGLVLDPKWDKATRQAKLQEVRLAKHFEEAKGGSDGALIYDLNLK
GLVAEVFGGGLGRLNQLNRPLSSASIGPADLLEVPPGGITLEGVRFNTEIVVRFIDSWLRGRGSFVYR
NSAEDSATAEISRSQIWQVVRHGLYTEGKDAITLSLVMFACETADELAQEGREAQIPALVDDRVS
ALSLYRMLVSVPVFPRFITTFLYEQALFHEFARRRSPIM

C. 3-Ketoacyl thiolase (NgPKT, EWM24705.1) :

MSSTQTSQANK**RLHLSGHL**VKK (intron 1) TSRPGFLGGDVVVVSALRTPL**C**KAKRGAFKSTT
TDDLAPVLEAVVKQSGVDAATLGDIVGNVLQPGSGAVGARMAQFYAGIPYQ (intron 2) VPL
CTLNRQCSSGLQAFQLQAASIQSGLYEAGIAGGVESMSLTDMSVDPVNFERSVENSALSKDCTIPMG
QTSDEVATRFQVSREDQDRFAAASHAKAEAAVKAGKFAEEIVPVRVSSGGEDGGEDVTVVREDEGIR
PGTTPEKLGLRSSFSFEGGSTTAGNSSQMTDGAAVLVMSRGAATAQGMFVMGVLGAAVGVPPDIM
GIGPAVAIPAALQARLRVEDVDVFEINEAFASQCLYCVRELKIPMEKVNPNNGGAIALGHPLGATGAR
QIATLLHEMRTRKRWGVVSMCIGTGMGAAVFNAAAAN

D. Histidine triad family protein 1 (NgHIT1, EWM29206.1) :

MSQ**RLVRLSQHL**LVHRTSLHATSIAVRLSRTTR**C**LVTGNQANTMADEVAESRKAAAKKMEEAADAGE
(intron 1) PTVFDKIVKKEIPSNFIYEDDDCVAFHDLSPQGPVHFLVIPKDRAGLSRLSKAEESH
KALLGHLLYVAQQVAKQEGVPGGFRVINDGPDGSQSVYHLHIHVIGGRQ (intron 2) MGWPPG

Suppl. Figure S6: Exon-intron structure of the PEX7 ortholog and three predicted PTS2 proteins from *N. gaditana* investigated experimentally in this study. For each *N. gaditana* protein, the Genbank accession number is provided for strain B-31 and intron positions are indicated. For the three PTS2 proteins investigated experimentally, the CDS of the first exon (underlined) was cloned from strain CCMP526 and placed upstream of the fluorescent reporter gene. The predicted PTS2 are highlighted in red and the most conserved four amino acid residues (e.g. RLX₅HL) are marked in bold. Cys residues of putative PTS2 cleavage sites are indicated (blue, C, D).

Suppl. Figure S7

P_inf	172	-----ETKAPVPMHKPMQAAFPVAAQQADTLDLLEAAKNLEAQQ--ASSEMART
P_tri	148	-----
E_sil	221	-----EK-----GVGEGQVGASARSL-VGQ
A_tha	267	VNGWATEFEQGQSOLMSSQMRS-----MDM-----QNIAAAME-QTRK-L-AHT
G_the	83	-----EAWAEFR-----ERE-HGSE-L-SRA
C_ele	158	-----
H_sap	209	LQHTA-----SDF
N_oce_N-term	252	WSHAGQAF-----PALEESH-----DA-----QATRAASGAVLEALREGG
N_oce_C-term	2	-----
N_oce_full	252	WSHAGQAF-----PALEESRVEAGSSSLPPSLEASLPPSLHEDDA-----QATRAASGAVLEALREGG
N_gad	1	-----
consensus	351	-----
P_inf	219	MS--QNPISKFNQSCFKEFNQNSACEVQITEKNEVINGHLKMEGALEGAMEDTSD-MHSNRDL---F
P_tri	148	---QMARVMVQQQLMTRAQNMSEHHHHRIRESNEQKSVSLDN-----QONGVD-KEFGQ
E_sil	240	MA--ADPLGGRFRLSELRLRATRTGTGGVRVSGDK--VTPGSGEAQG-LDEAMAGGASSQQHQQAAGDPAV
A_tha	307	LSQ--DGNPKFQNSRFQFMSKSRGELIIEENQ--VKQASAPGE---MATE---YEQNYL---G
G_the	103	LEETLNGIEKLAQSELRLDFHAAVNGSGFQFNGS--LQPDGE-----
C_ele	158	-----
H_sap	217	VA--KVDIEPLANSEFFKFWROVGEQMSLSGA--GSGRA---QA-----
N_oce_N-term	287	RE--GGVCAKMARSEFFGFFSQVNKGEIAFEGNT--V-----CTTAE-ERRHQAQGLGL
N_oce_C-term	2	---ARSEFFGFFSQVNKGEIAFEGNT--VPRAGDFEG--AAETAE-ERRHQAQGLGL
N_oce_full	310	RE--GGVCAKMARSEFFGFFSQVNKGEIAFEGNT--VPRAGDFEG--AAETAE-ERRHQAQGLGL
N_gad	1	-----
consensus	421	k sefl fm qv g v d vv w q
P_inf	282	D-----ASWKQS-----EN-----G--SA
P_tri	200	-----VS
E_sil	305	AAAATKADFQVAYAEN--EAAQQQQQGP-----LVGSFSDAWNGLDSGVGVREEPSSL-KPSAAEILLS
A_tha	359	PP-----SWADQFANEKLSHGP-----
G_the	144	-----
C_ele	158	-----N--WADFMEQQ-----
H_sap	256	-----EQ--WAAFIQQQ-----
N_oce_N-term	338	EG-----EW--EAAARAAVE-----
N_oce_C-term	54	EG-----EW--EAAARAAVEGPDIGAAAGLEGAWAEA-----HRVRGGQGADLSGS--VK
N_oce_full	371	EG-----EW--EAAARAAVEGPDIGAAAGLEGAWAEA-----HRVRGGQGADLSGS--VK
N_gad	1	-----
consensus	491	a a e
P_inf	294	AA--ME--NMF-SEASAA-H-A---HPHGAWKEAGTANA-----TSLDQAWG-----SK
P_tri	202	DM--AP--VMHEGVTTQG-VSM---EEAANAWEVADDNITV--G-H-----EG
E_sil	365	SA--QE--EAW-SDGKTG-L-APDFQEKMAATREVEQGGVGAGVGDPDLQAIWEESDD---DAAG-
A_tha	376	-EQWAEFASR-GQQETA-----BDQNV--NEFSK-----LNVD-DWI-----EF
G_the	144	-----
C_ele	168	-----DN-YG-----MENTWK-----
H_sap	267	-----G-----TSDAVV-----DQFTRPVNTSALDMFE
N_oce_N-term	-----	-----
N_oce_C-term	100	AG--LD--EAW-AQAEVEAK-N---ADHSLNKEGAEGKD-----LDMDAFWDHVQSVAGEYNA
N_oce_full	417	AG--LD--EAW-AQAEVEAK-N---ADHSLNKEGAEGKD-----LDMDAFWDHVQSVAGEYNA
N_gad	1	-----
consensus	561	a le w d
P_inf	335	TAEKMMSAWGESD-N--EETHEKAMAAQ--TTDPFEDAWDNAT--NQDITKAEVPSDSSNFKG
P_tri	240	-----LAQGATIEE--EAAQAQAEAYDSVDAATNLWNTDNDP--VYELNTEKPERDQQWMEQG
E_sil	423	-----V-ETDGVVSRTAATLE--A---GEGNL-----EAPFELSABNRNVDSPFEG
A_tha	413	AEGFVGDN--S-----SADAMANAYDFLNEKNAGK---QTS-----GVVVSIMNPPVGHFPMKEG
G_the	144	-----LTD-----PAGL---ETPF--SLRPFITTPHNPFGRPFCEKRS
C_ele	178	-----DAQA--FEQWEEIKRMEKDESQ-----SPENVVYQANPPTTMSPLMEG
H_sap	291	RAKSAIE--SDVDWFDK--EAELEEMAKRDAEAHPWLSDYDDLTSATYDKGQOEENPLRCHPQFEG
N_oce_N-term	-----	-----
N_oce_C-term	151	MAKDAVPGISKADAV-E--EESWQQQARLEEWAARGAAEGGLEGG--QAAGGADDNEFNDHAAABEG
N_oce_full	468	MAKDAVPGISKADAV-E--EESWQQQARLEEWAARGAAEGGLEGG--QAAGGADDNEFNDHAAABEG
N_gad	1	-----
consensus	631	l a w e y y e n fld e f e g

Suppl. Figure S7: Analysis of *Nannochloropsis* PEX5 orthologs for the presence of the PEX5 binding domain in Stramenopiles. (Legend on next page).

Suppl. Figure S7 (cont.)

```

P_inf      399  IEFFKSGHIDDAI LAFEAEEVQQ-HENSEAWRMLGECHAENDEDKSAITCLERAVEBDFPYNLSALLALGV
P_tri      299  IREFNAGNKEBAKAFETIEIYVYCNNGNSAAWKMLGRCHAENDMDREBAIVCLEQAVDRDPKSPPEALLLGV
E_sil      467  VRIFFEGQADAAICFFAEIAR-NFENSQAWFVLGQSHAENDQDRLAISCLEKAVETIDPYSLLALLALGT
A_tha      465  QEIRKRLLSBAALAEAEVMK-NFENAEAWRMLGVTHAENDDQQAIAAEMRACADPTNLVLLALGV
G_the      178  VEYLQRCHEVESGALAEAEVQE-HEDNCLAWLTLGLAHAENDEIVKAIIAINRAVQADPDLNALLALGV
C_ele      224  DNIMRNGDGNAMLAIEAAVQK-DEQDARAWCKLGLAHAENKQQLAQAFQKCIQTAGNKBAALLLSV
H_sap      358  IRLRLQRCGDPNVAIILFEAAEVQQ-DEKHMELAWQVLTTCBAENQQLLAISAIRCELELKEEDNQALMALAV
N_oce N-term
N_oce C-term  217  VAIIEACEIRCAI LAFEAEEVQR-AEHNAAEWRLMLGTHAENDEDKRAIACLERAVTHDPHSLPTLLALGV
N_oce full    534  VAIIEACEIRCAI LAFEAEEVQR-AEHNAAEWRLMLGTHAENDEDKRAIACLERAVTHDPHSLPTLLALGV
N_gad        1  -----AENDEDKRAIACLERAVTDPHSLPTLLALGV
consensus    701  i l f g l ailafeaevg p n eaw mlg haendedk ai clerave dp l allalgv

```

```

P_inf      468  SNVNEINPQGALKTKLAWIHNPKYHGLEIIVDEY----SDG-----SIMDEVVMOLML
P_tri      369  SYVNEINNAKAKIKLKAWEIHNPKYAGMEIVQVLMY----RDS-----LVDQESAFDEVORLLV
E_sil      536  SYVNEILDSQKALTILKAWIEHNPKYSGLEIIVDEY----SDG-----SIMDEVVMOLML
A_tha      534  SHNNELEQATALKYLYGWNRNHPKYGA-LAPEE----LADS-----LYHADLARLFN
G_the      247  SHNNELEQVNALITLISMTIRHEFISCLCPPEHIGSLRETY-----LHNEVITQFT
C_ele      293  SQANEGMENEALHOLDRWVSSYLGSNSTQITTTTP-----LY-----SSFDS-DTFNRVREARFL
H_sap      427  SFNNEILQRCACETLIRWRYRFAAHIVTPAEIG----AGGAGLGPSPKRILGSLSD-SIFLEVKELFL
N_oce N-term
N_oce C-term  286  SYVNEILDSVRALQNLKAWIEHNPKYQGLEIIRVDEY----SDG-----SIMDEVVMOLML
N_oce full    603  SYVNEILDSVRALQNLKAWIEHNPKYQGLEIIRVDEY----SDG-----SIMDEVVMOLML
N_gad        33  SYVNEILDSVRALQNLKAWIEHNPKYQGLEIIRVDEY----SDG-----SIMDEVVMOLML
consensus    771  synneild al nlkaw hnpky glei vdey dg sl dev qlml

```

```

P_inf      517  QARAHLP-----S-LS-DVQVVLGVLYNVSDYDAVVSSEFAATDSCQEDHYALWNKLGATLANS
P_tri      423  QALEYLP-----V--DASVLEAGVLYNVSDYVAAGGAFFRAIDAREDDYQLWNKLGATLANGN
E_sil      585  QAQRMLA-----T--DA-DAHVLGVLYNVSDYLSAAEAFFRAIEARPNIHSLWNKLGATLANSR
A_tha      581  EAQSQLPN-----E--DA-DVHVLGVLYNLSREHFRATTSQQTALQLPNDYSLWNKLGATQANSV
G_the      300  RVLQTRP-----D--DV-EHTVLGLYLYHISVDYDKAIEHFREAIRINPDYSLWNKLGATLANSF
C_ele      347  DAARQOG-----ATPEP-DQNALGVLYNINENARAVDSLRDLASKNPTFARLWNKLGATLANGD
H_sap      492  AAVRLLP-----TSTLP-DVQCGLGVLYNLSGEYDKAVDCEFAALSVRPNDYSLWNKLGATLANGN
N_oce N-term
N_oce C-term  335  QAHDWQATSGPGGQGS--DA-DVQVVLGVLYNVSDHDESAATAAFRAALLSRPSDYSLWNKLGATLANSQ
N_oce full    652  QAHDWQATSGPGGQGS--DA-DVQVVLGVLYNVSDHDESAATAAFRAALLSRPSDYSLWNKLGATLANSQ
N_gad        82  QAQDWMRATCGPEG--Q--DA-DVQVVLGVLYNVSDYDLSAAAFAFRAALVAPRPSDYSLWNKLGATLANSQ
consensus    841  qa d da dv vvlgvlynvsdkdyd a fr al rp dyslwnklgatlans

```

```

P_inf      575  RSSEAPAYHRALELKYARGWNLGISANLGNYPEATKCYLCALSLNNR---A-----MHI
P_tri      482  QSEAPAYHRALELKEKYARGWNLNMTSHSNLQNYDEAAKCYLOTSLNFA---A-----MHC
E_sil      643  QSEAAPAYHRALEKKEYARGWNLNMGISANLNRYEASSCYLCALRLNPE---A-----MHI
A_tha      639  QSAALSAVQCALDLKENYVRAWNLNGISANNGMYKESIPYVRAALNPK---A-----DNA
G_the      358  KSEAPDAVICALSKKENYVRLANLGTAYSNQEMYEEASCYIKALSNEG---A-----SHI
C_ele      407  HTAEASAVRPALELKYETVRRYRNGLISCMQSSYDEALHHSALDELKGGNDA---SGI
H_sap      552  QSEAPAYHRALELQEGYHRSRYNGLISCINLGAHREAVEHHTALNNQR---KSRGPRGEGGAMSENI
N_oce N-term
N_oce C-term  402  RSSEAPAYHRALELKEKYARGWNLNGISANLSSFDEAAKCYLCALRLNPE---A-----MHI
N_oce full    719  RSSEAPAYHRALELKEKYARGWNLNGISANLSSFDEAAKCYLCALRLNPE---A-----MHI
N_gad      147  RSSEAPAYHRALELKEKYARGWNLNGISANLGRYEEAAKCYLRAIRLNPE---A-----GHM
consensus    911  rs eaipayhralelky argwlnlgis anl yeea kcy l al lnp a hi

```

```

P_inf      631  WSYLRITCF--CMERFDLVKIADTKIAREREKFLIDL----
P_tri      538  WSYLRITAL--SCSERFDLVQHAASONLEAKDLDFEVIYS---
E_sil      699  WSYLRITVE--SMERFDLVQKAGKEAGLEEDDFDLALPTPPP
A_tha      695  WQYLRITSL--SCASRCDMEIACESRNLDLLQKHFPL-----
G_the      414  WSSLEISVF--HFVDRSDLAARCDTENAEARGIDF-----
C_ele      466  WITMRSAAIR--SNVPDNLRAVERRDLAAVKASLV-----
H_sap      619  WSYLRITAL--SMGQSDIAYGADARDLSTLLTMFGLPQ-----
N_oce N-term
N_oce C-term  458  WGYLRITF--SMERFDLVQLAGQQDPSVEDAEFGR-----
N_oce full    775  WGYLRITF--SMERFDLVQLAGQQDPSVEDAEFGR-----
N_gad      203  WGYLRITF--SMERFDLVQLAGQQDPSLEAEAFGR-----
consensus    981  wsylri f t merfdlv a d f ef

```

Suppl. Figure S7: Analysis of *Nannochloropsis* PEX5 orthologs for the presence of the PEX7 binding domain in Stramenopiles. The predicted genomic PEX5 CDS from *N. gaditana* (N_gad, EWM20982.1) and *N. oceanica* (626268) were aligned with PEX5 homologs from *A. thaliana* (A_tha, NP_200440), *H. sapiens* (H_sap, NP_001124495) and others by ClustalW (v2.1, at Phylogeny.fr). Sequence conservation was labelled by Boxshade. The alignment confirms that the four *Nannochloropsis* proteins are indeed PEX5 homologs and possess the conserved middle domain (boxed in red) that binds PEX7 in *Arabidopsis* and humans. The alignment of the weakly conserved N-terminal domain (approx. 260 aa of *A. thaliana* PEX5) is not shown.

Suppl. Figure S8

N_oce_mRNA 1 MPRPGCFPPFPGGPPFGGHGKYPGEVPDAVFHTQ-FQCCSVFEFSPFHEGRRLAVATAQYFGIIGNGRONI
N_gad 1 MKRESFHAANTGGGGLGG-----PIPDADFQTQ-FQCCSVFEFSPFHEARLAVATAQYFGIIGNGRQHI
A_tha 1 MF-----VFRAP-FNGYSVKFSPFYESRLAVATAQNFGIIGNGRITHV
H_sap 1 MSAVCGGAAR-----MLRTPGRHGYYAEFSPLYLRGLRLACATACHVGIAAGC--TL
consensus 1 * * * * *

N_oce_mRNA 70 LEMGPDGH-LRESRNFLTCEGLYDCSWSEVNENOLVSASADGSLKLWDVTMS-DDYPIAHWHEHQAEASS
N_gad 63 LETGPDGN-LREIRSFLTCEGLYDCSWSEANONOLVSASADGSLKLWDVTMS-DGYFAHWQEHSAEVSS
A_tha 42 LELAPGPGVETESVSDTDAVDVCWSSEHSVLTAIGDGSVKLYDTALPPPSNPTRSRFOEHAREVCS
H_sap 49 LIIDPDEAGLLERSFDNDNLGDVTVSENNEHVLIITCSGDGSLQLWDVAKA--AGPLQVWEKEHAQEVYS
consensus 71 * * * * *

N_oce_mRNA 138 VDWNLVCKNTFLSS-SWDGSIKHWDPTRPSSLATYMLGHDGS CVYNNAVWSPRHATRFILSCAADGSVRVWDA
N_gad 131 VHWNVQVKTNFLSA-SWDGSIKLWDPHPHPTLSITYCGHTG-CVYAGIHSPRHHRFRLSCGTGDSLRIWDT
A_tha 112 VDDVVTKLWMDRFPASVRTPEKRAY-CVYQAMWNPKHGDVEFASAGDQDLRIWDV
H_sap 117 VDWSOTRGELVSGSWDCTVKLWDPTVGKSLCTFRGHES-ILYSTIWSPHHGCFFASAGDQDLRIWDV
consensus 141 * * * * *

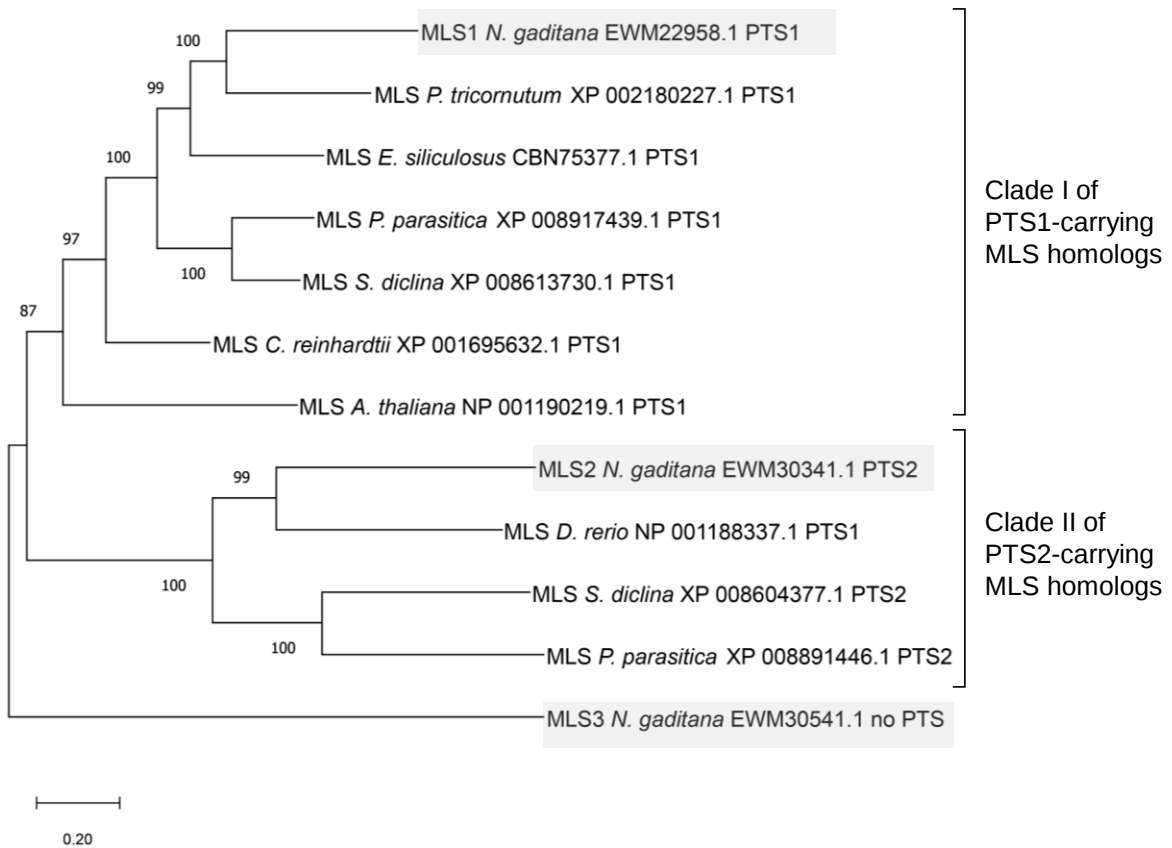
N_oce_mRNA 207 SLPPSLPPAMGLGGGMEGMVQVVRRAHEGEVLADWDKYNEWVYSGGVDRSVKIWDLRPSLPPLSLFLH
N_gad 199 KLPPSHATSILGLG--RAEGGAVQVVRRAHEGEVLSADWDKYQDFLVYTGGVDRSKIWIHLRPSLPPLGLH
A_tha 180 REP-----GSTMTIPAHDFEILLSCDWNKYDCHCATSSVDKTIVKVWVRSYRVPLAVLN
H_sap 186 KAA-----GVRIVIPAHDFEILLSCDWNKYENLVITGAVDCSLRGWDLNRVROPVFEILL
consensus 211 * * * *

N_oce_mRNA 277 GHGFAVRRLKASPHOCELIGSYSDMSFRLWGAAREGERVGVVWRCEYHTEFVQGLEDLFMPGRIATCS
N_gad 267 GHGYAVRRLLKTSPOEGVLGSVSYDMSFRLWGPTFRAGRAGELWRCEHTEFVQGLDHFLEMPGRIATCG
A_tha 234 GHGYAVRRVKFSPHRRSLIASCYSDMSVCLWDYMVEDLVG--RYDHHTEFVAGIDMSVLEVGLMASTG
H_sap 240 GHTYAIRRVKFSPFHASVLAISCYDYFVRFNNFSKPDSLET--VEHHTEFTCGLDPSLOSFTQVADS
consensus 281 * * * * *

N_oce_mRNA 347 WDRRCVCRMRPFI----
N_gad 337 WDRRCVCRMTLPIPI----
A_tha 301 WDELIVYVWQQGSDPRAS
H_sap 307 WDETIKIYDPACTIPTIPA
consensus 351 * *

Suppl. Figure S8: Homology analysis of two putative PEX7 homologs from *Nannochloropsis* by multiple sequence alignment with *Arabidopsis* and human PEX7. The translated CDS of a full-length mRNA of a putative PEX7 homolog from *N. oceanica* CCMP1779 (N_oce_mRNA, 583661) was aligned with the predicted PEX7 from *N. gaditana* (N_gad, EWM28214.1) and with PEX7 from *A. thaliana* (A_tha, NP_174220) and *H. sapiens* (H_sap, NP_000279) by ClustalW (v2.1, at Phylogeny.fr). Sequence conservation was labelled by Boxshade. The alignment confirms that both *Nannochloropsis* proteins are indeed PEX7 homologs of *Arabidopsis* and human PEX7 and that *N. oceanica* expresses a full-length PEX7 mRNA.

Suppl. Figure S9



Suppl. Figure S9: Phylogenetic analysis of peroxisomal MLS isoforms from *N. gaditana*. Homologs of the PTS1- and PTS2-carrying MLS isoforms of *N. gaditana* (NgMLS1, EWM22958.1, SRL>; NgMLS2, EWM30341.1, RLx₅HL) were identified by protein BLAST searches at NCBI using Genbank and focusing on Stramenopiles, green and red algae, land plants and animals. Multiple sequence alignment of homologous proteins was performed in MUSCLE (Edgar 2004) and the proteins were analyzed for predicted PTS1 using the PredPlantPTS1 prediction server (<http://ppp.gobics.de>) and for predicted PTS2 using a manual motif search algorithm. The phylogenetic tree was constructed using the Bayesian inference method (Ronquist 2012). The branch support values were calculated as Bayesian posterior probabilities and are shown next to the branches. The platform Phylogeny.fr (Dereeper et al., 2008) was used for phylogenetic analysis, and the tree was visualized in MEGA X (Kumar, 2018). Among selected organisms, only Nannochloropsis and two oomycete species (*Saprolegnia diclina* VS20 and *Phytophthora parasitica*) had both, a predicted PTS1- and a PTS2-carrying MLS isoforms. The phylogenetic analysis revealed two distinct clades of PTS1-carrying and PTS2-carrying MLS isoforms (except for *D. rerio*). The non-peroxisomal MLS isoform of *N. gaditana* was used as an out-group.