Diversity and evolution of cytochromes P450 in stramenopiles

Linhong Teng, Xiao Fan, David R. Nelson, Wentao Han, Xiao Zhang, Dong Xu, Hugues Renault, Gabriel V Markov, Naihao Ye

Diversity and evolution of cytochromes P450 in Stramenopiles

Linhong Teng[1,2], Xiao Fan[1], David R. Nelson[3], Wentao Han[1], Xiaowen Zhang[1], Dong Xu[1], Hugues Renault[4], Gabriel V. Markov[5], Naihao Ye[1,2,*]

[1]Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, China, 266071
[2]Function Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, China, 266071
[3]Department of Microbiology, Immunology and Biochemistry, University of Tennessee Health Science Center, 858 Madison Ave. Suite G01, Memphis, TN, USA 38163
[4]Institute of Plant Molecular Biology, CNRS, University of Strasbourg, 67084 Strasbourg, France
[5]Sorbonne Université, CNRS, Integrative Biology of Marine Models (LBI2M), Station Biologique de Roscoff (SBR), 29680 Roscoff, France

*Author for correspondence: Naihao Ye, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, China, Tel/Fax: +86-532-85830360, E-mail: yenh@ysfri.ac.cn;

Main conclusion

Comparative genomic analysis of cytochromes P450 revealed high diversification and dynamic changes in stramenopiles, associated with transcriptional responsiveness to various environment stimuli.

Abstract

Comparative genomic and molecular evolution approaches were used to characterize P450 diversity in stramenopiles. Phylogenetic analysis further supported the high diversity of P450 in stramenopiles. Three clans were defined, that is, CYP51 clan and CYP97 clan present in brown algae, diatoms and *Nannochloropsis gaditana*, CYP5014 clan present in oomycetes. Two ancient families, CYP51 and CYP97 showed a close relationship with red algae, green algae and higher plants. Gene gain and loss patterns revealed that six CYP families – CYP51, CYP97, CYP5160, CYP5021, CYP5022, and CYP5165 – predated the split of brown algae and diatoms. After they diverged, diatoms gained more CYP families, especially in the cold-adapted species *Fragilariopsis cylindrus*, in which eight new CYP families were found. Selection analysis revealed that the expanded CYP51 family in the brown alga *Cladosiphon okamuranus* exhibited a more relaxed selection constraint compared with those of other brown algae and diatoms. Our RNA-seq data further evidenced that most of *P450s* in *Saccharina japonica* are highly expressed in large sporophytes, which could potentially promote the large kelp formation in this life stage. Analysis of *Ectocarpus siliculosus* and diatom microarray transcriptomes showed that many *P450s* are responsive to stress, limited nutrient or light quality, suggesting their pivotal roles in detoxification or metabolic processes under adverse environmental stimuli. The information provided in this study will be helpful in designing functional experiments and interpreting P450 roles in this particular lineage.

Keywords: Cytochrome P450; Comparative genomics; Molecular evolution; Stramenopiles

Introduction

Cytochromes P450 (P450, CYP) are monooxygenase enzymes found in all life kingdoms, from prokaryotes to eukaryotes, which catalyze a broad range of irreversible, often rate-limiting reactions within metabolic pathways (Nelson 2017). They are one of the largest enzyme superfamilies, showing singular patterns of evolution, including diversification bursts (Nelson 2017). All extant P450s were known to originate from a common ancestral gene, and evolved into diverse forms having different activities (Omura and Gotoh 2017). P450s are involved in the biosynthesis of important specialized metabolites (e.g. sterols, pigments, antioxidants), which are thought to convey adaptive advantages in specific ecological niches. Furthermore, they participate in the detoxification of xenobiotics such as pesticides (Yu et al. 2015). Considering their catalytic diversity, CYPs also offer prospect for potential

applications in medicine, agriculture and industry through metabolic engineering approaches (Syed et al. 2013; Renault et al. 2014).

The comprehensive repertory of *CYP* genes in the genome of a defined species is called the CYPome. Nowadays, with the breadth of sequencing projects, more than 350,000 P450 sequences are available, over 41,000 of which are named, although the functions of the majority are still unknown (Nelson 2017). Nomenclature of CYPs is based on amino acid sequence identity. Generally, sequences having more than 40% identity belong to one family, and sequences having more than 55% identity belong to one subfamily. CYP families ranging from 1 to 49, 301 to 499 and from 3001 to 4999 are for animals, 51 to 69, 501 to 699 and 5001 to 6999 for lower eukaryotes, 71 to 99, 701 to 999 and 7001 to 9999 for plants, and 101 to 299 and 1001 to 2999 for bacteria. (Nelson 1998; Phillips and Shephard 1998). Ultimately, family assignment is validated by phylogenetic reconstructions. CYP clans are higher classification units. They are deep branching clades on phylogenetic trees that form natural divisions among CYP sequences from one kingdom. Animals have 12 clans and multicellular land plants have 12 too. Clans for algae are not well defined yet (Nelson 2017).

Stramenopiles (also referred to as heterokonts), together with alveolates and rhizarians are called 'SAR clade'(Keeling 2013). The ancestral plastid in this group was derived from red algae via secondary or higher endosymbiosis events, in an organism that may have possessed a green algal plastid (Frommolt et al. 2008; Moustafa et al. 2009) ; Dorrell et al., 2017). The photosynthetic members of the stramenopiles, termed the ochrophytes, are the most diverse algal group with plastids of algal origin (Dorrell et al. 2017). The best characterized ochrophytes, diatoms and multicellular brown algae, constitute two important groups of stramenopiles (Ševčíková et al. 2015; Aleoshin et al. 2016). Diatoms are major primary producers in the ocean (Mock et al. 2017), while submarine kelp forests are the largest marine communities, which serve as spawning grounds for marine animals, as well as being important components for future renewable energy resources (Ye et al. 2015; Teng et al. 2017a; Teng et al. 2017b). Stramenopiles also contain non-photosynthetic lineages, such as oomycetes, which behave as pathogens for many organisms, including humans (Dorrell et al. 2017).

The first evidence of P450 in marine macroalgae, including green algae, brown algae and red algae was obtained by CO-difference absorption spectra, though at rather low levels (Pflugmacher and Sandermann 1998). Active CYPs in marine macroalgae could remove marine xenobiotics by metabolizing several standard P450 substrates, which extended the green liver concept to these species (Pflugmacher and Sandermann 1998). The first identified stramenopile P450 sequence, *CYP97E1*, was cloned in the diatom *Skeletonema costatum* (Yang et al. 2003). Brown algae produce a vast array of specialized metabolites, such as diverse oxylipins (Kousaka et al. 2003; Ritter et al. 2008). The recently cloned CYP5164B1 in the brown alga *Ectocarpus siliculosus* converts 9 and 13-hydroperoxides of linoleic acid into oxylipin epoxyalcohols (Toporkova et al. 2017). Besides, blooming of CYPs in oomycetes has been reported with respect to their diversity and duplications, suggesting a role in adaptation of oomycetes to diverse ecological niches (Sello et al. 2015). However, stramenopile P450 diversity remains relatively unexplored, especially in the recently sequenced brown algae and diatoms. Taking into account the great importance of CYPs in metabolism and the paucity of knowledge about them in marine stramenopiles, this work performed systematic identification and comparison of P450s within representative lineages of stramenopiles to provide a detailed characterization and evolutionary history of the CYPome in this particular lineage.

Materials and methods

Data source

The most up-to-date genomes of the three brown algae (*Ectoparpus siliculosus*, *Cladosiphon okamuranus* and *Saccharina japonica*), the three diatoms (*Phaeodactylum tricornutum*, *Thalassiosira pseudonana* and *Fragilariopsis cylindrus*) and the *Nannochloropsis gaditana* are available from public databases. Briefly, genome sequences of *E. siliculosus* V2016 version were downloaded from the website http://bioinformatics.psb.ugent.be/orcae/overview/Ectsi (Cormier et al. 2016). Sequences of *C. okamuranus* were downloaded from http://marinegenomics.oist.jp/algae/ (Nishitsuji et al. 2016). Sequences of *S. japonica* were downloaded from NCBI https://www.ncbi.nlm.nih.gov/. Sequences of *F. cylindrus* were downloaded from JGI https://genome.jgi.doe.gov/portal (Grigoriev et al. 2014). Sequences of *P. tricornutum*, *T. pseudonana*, and *N. gaditana* were downloaded from ENSEMBL http://protists.ensembl.org/info/website/ftp/index.html (Kersey et al. 2018). Besides, the publicly available P450 sequences from two oomycetes *Phytophthora sojae* and *Phytophthora ramorum*, and other species (CYP51 and CYP97 from animal, plants, fungi, bacteria) were retrieved from the P450 homepage http://drnelson.uthsc.edu/CytochromeP450.html (Nelson 2009). Add ref here to transcriptomic datasets.

Identification of P450s

The P450 genes within stramenopiles were identified as described below. First, the P450 domain PF00067 was downloaded from the Pfam website http://pfam.xfam.org/ (Finn et al. 2016). HMMER3 http://hmmer.org/ (Eddy 2011) software with the default parameters was used to search for PF00067 domain in the proteomes of each species including the three brown algae, three diatoms and *N. gaditana*. Then, we used acquired proteins as query sequences for tBlastn and Blastp searches (Lobo 2012) against the genome sequences and predicted proteome sequences, respectively. Redundant sequences were discarded from the data set to obtain unique P450 genes. All of the obtained P450s were further manually examined using the NCBI online BLAST tool https://blast.ncbi.nlm.nih.gov/Blast.cgi to confirm the genes' annotation. The online InterProScan program http://www.ebi.ac.uk/interpro/search/sequence-search was used to confirm the presence of P450 domain in each sequence. All the putative P450s were named by David R. Nelson based on the existing nomenclature criteria, that is, > 40% sequence identity for assigning a family, > 55% for a subfamily and < 40% for a new family. Family assignment was further confirmed by phylogenetic reconstructions.

Sequence structure analysis

For each lineage (i.e. oomycetes, brown algae and diatoms), sequence structural features were explored to reveal a lineage-specific conservation pattern. Multiple protein sequence alignments were obtained using the hmmalign program in HMMER 3.1 package, with CYP protein sequences and PF00067.hmm as input files. Sequences logos of conserved motifs were generated using the online Weblogo program http://weblogo.threeplusone.com/create.cgi (Crooks et al. 2004) based on the sequence alignment results. Intron and exon information of the CYPs was obtained from their GFF3 files and was used for the graphic display using the Gene Structure Display Server of Peking University http://gsds.cbi.pku.edu.cn (Hu et al. 2014). Subcellular protein localization was predicted using ASAFind http://rocaplab.ocean.washington.edu/tools/asafind (Gruber et al.

2015) and Euk-mPLoc 2.0 http://www.csbio.sjtu.edu.cn/bioinf/euk-multi-2/ (Chou and Shen 2010). The protein transmembrane helices were predicted by TMHMM server 2.0 http://www.cbs.dtu.dk/services/TMHMM/ . The distribution of P450s on chromosome or scaffold was displayed using Mapchart (Voorrips 2002). Genes separated by 0-5 genes were suggested to be tandem duplicates (Yu et al. 2015; Zhao et al. 2012).

Phylogenetic analysis

P450 protein sequences in stramenopiles were extracted and aligned using CLUSTAL Omega http://www.ebi.ac.uk/Tools/msa/clustalo/ (Sievers et al. 2011). The alignment file was used as input for MEGA 7.0 (Kumar et al. 2016). Maximum likelihood (ML) phylogenetic trees were constructed with the best substitution models Le-Gascuel (LG) + gamma (G) + Freqs. (F) model, as predicted by the menu 'Find best DNA/protein models' in MEGA 7.0. Bootstrap with 1000 replicates was performed to obtain the confidence support level. Besides, the phylogenetic relationships among the of CYP51 and CYP97 families, including stramenopiles, plants (red algae, green algae and higher plants), animals, bacteria and fungi, were reconstructed using protein sequences to investigate the phylogeny origin of these two global CYP families in stramenopiles.

Evolutionary history analysis

The ratio of non-synonymous (d*N*) to synonymous (d*S*) nucleotide substitution rates (d*N*/d*S* = ω) provides information about the evolutionary forces operating on a gene. As positive selection promotes non-synonymous substitutions, an $\omega > 1$ is considered indicative of positive selection, while genes under negative selection show an $\omega < 1$. When genes are kept under neutral evolution, rates of synonymous and non-synonymous substitution are equal and $\omega = 1$ (Teng et al. 2017a). Thus, we estimated the selection pressure direction and strength of each CYP family by assessing ω values. The coding sequences of CYPs were translated into amino acid sequences using a local perl script. Alignment of these amino acid sequences was performed using CLUSTAL Omega http://www.ebi.ac.uk/Tools/msa/clustalo/ (Sievers et al. 2011). Codon alignments were generated using the protein sequence alignments as a guide (Suyama et al. 2006). The corresponding gene tree of each family was generated using MEGA 7.0. Then we used one ratio model (M = 0) of Codeml program in the PAML package (Yang 2007) to calculate ω within each CYP family. Besides, site model, free ratio model, branch specific model and branch-site model were used to detect positively selected genes. To elucidate the evolutionary process of CYP families, we performed family history reconstruction using the COUNT software with default parameters (Csűös 2010). This software implements phylogenetic birth-and-death model, in which family size evolution on an edge is governed by a linear birth-and-death model. Ancestral reconstruction can be carried out by computing posterior probabilities for the family sizes at inner nodes. The species tree including the three brown algae and three diatoms, together with the numerical gene profiles (number of genes present in each family per species) were input into COUNT software and the data were analyzed using the Dollo parsimony model.

Expression profile analysis

Previous published microarray data from *E. siliculosus* were used to explore the expression levels of P450s in response to abiotic stresses, including hyposaline stress, hypersaline stress, oxidative stress, and Cu stress (Dittami et al. 2009; Ritter et al. 2014). The expression level was determined by averaging the expression values (previously quantile normalized by Roche

NimbleGen, Madison, WI, USA) of four replicates for each experimental condition. Furthermore, expression levels of *S. japonica* P450s were examined in different life cycle stages including sporophytes, male gametophytes, and female gametophytes. Their transcriptomes were sequenced by our group using Illumina 2500, with three biological replicates in each life stage. The SRA accession numbers in the NCBI database are SRR5860560, SRR5860561, SRR5860562, SRR5860563, SRR5860564, SRR5860565, SRR5860566, SRR5860567, and SRR5860568. Gene expression levels were estimated by FPKM (Fragments Per Kilobase of transcript per Million fragments mapped). Differential expression analysis among life stages was performed using the DESeq (Love et al. 2014) R package using a model based on the negative binominal distribution. The resulting P values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate (Benjamini and Hochberg 1995). Genes with an adjusted P value < 0.05 found by DESeq were assigned as differentially expressed. Transcriptome-wide microarray expression data of diatoms were acquired from the pan-transcriptomic analysis (Ashworth et al. 2016) http://networks.systemsbiology.net/diatom-portal/. The integrated data used in this analysis included: silica, iron, and nitrogen limitation, low temperature and elevated pH (Mock et al. 2008), exposure to pollutant benzo[a]pyrene (Carvalho et al. 2011), silica limitation (Sapriel et al. 2009), high light (Nymark et al. 2009), exposure to cadmium (Brembu et al. 2011), exposure to red, blue and green light (Valle et al. 2014). The microarray data were converted to log2 expression ratios *vs.* within-experiment control samples.

Results and discussion

Identified CYPomes in Stramenopile species

Genome-wide identification of P450s in three brown algae, three diatoms and *N. gaditana* revealed discrete gene numbers in each species of the three lineages (Fig. 1, Table S1, Fig. S1). The P450 count in brown algae selected in this study ranged from 12 in *E. siliculosus* to 22 in *C. okamuranus*. The diatoms, on the other hand, showed the highest P450 number of 23 in *F. cylindrus*. Partial sequences, possibly pseudogenes, were also identified in these species, including one in *P. tricornutum*, two in *S. japonica*, three in *C. okamuranus*, seven in *F. cylindrus*, and one in *N. gaditana*. When naming partial sequences, we added a letter P after the normal nomenclature. Annotation of P450 families and subfamilies revealed variable distribution within the stramenopiles. In total, three brown algae and three diatoms have 25 families as defined in the standardized nomenclature for P450 sequences. Comparative analysis revealed that the three diatoms have higher P450 diversity than the three brown algae and the two oomycetes according to the number of families, whereas the two oomycetes showed the lowest diversity (Fig. 1). The three brown algae share six families, the three diatoms share three families, and all the brown algae, diatoms and *N. gaditana* share only two families, whereas they have none in common with the oomycetes *P. ramorum* and *P. sojae*. As oomycetes P450s have been already reported (Sello et al. 2015), we focused in this study on brown algae and diatoms.

CYP51 and CYP97 catalyze essential reactions conserved in the plant kingdom (Mizutani and Ohta 2010). Both of them were present in every alga examined (Fig. 1a). All of the CYP51 belong to subfamily CYP51C by definition. Notably, the brown alga *C. okamuranus* has an expanded CYP51 family, whereas other species contain only one copy of CYP51 in each genome. Another conserved family across species is CYP97. Most species contain two CYP97 members, with one belonging to CYP97E and the other CYP97F. *F. cylindrus* has one

more CYP97F. *P. tricornutum* has two more CYP97E, with one CYP97E are partial sequence. Another shared family between brown algae, diatoms and *N. gaditana* is CYP5160. One CYP5160 exists in the three brown algae, *P. tricornutum* and *N. gaditana* whereas it is absent in *T. pseudonana* and *F. cylindrus*. CYP5161, CYP5163, CYP5164 and CYP5162 only occur in brown algae, of which the first three have several members per genome, whereas the later one CYP5162 features a single member. Among diatom-specific families, a single CYP5021 exists in all the three diatoms. CYP5022 and CYP5023 exist in *T. pseudonana* and *F. cylindrus*, while CYP5165 exists in *P. tricornutum* and *F. cylindrus*. In order to examine whether these families are indeed lineage-specific, we explored the transcriptomic database MMETSP (Keeling et al. 2014) and 1KP (https://db.cngb.org/onekp) to see the CYP distribution in related species. In close stramenopile relatives (*Bolidomonas* pacifica (?) close to diatoms and *Vaucheria litorea (?)* close to brown algae), identified CYP did not cluster with any lineage-specific CYPs (Fig. S2). Performing BLAST against 1KP database further supported that the brown algae-specific CYP families were only found in brown algae, CYP5023 was not found in the database, whereas diatoms' CYP5021, CYP5022 and CYP5165 were also found in other lineages, such as Haptophytes, suggesting they are not restricted to diatoms (Table S2). Notably, eight new families were found in *F. cylindrus*, that is, CYP5713, CYP5714, CYP5715, CYP5716, CYP5717, CYP5718, CYP5719, and CYP5720, suggesting their potential roles in adaptation to particular habitats. Indeed, *F. cylindrus* is a cold-adapted diatom, which lives in the extreme environment of the Southern Ocean, where photosynthesis is limited by large seasonal fluctuations in light, temperature and the extent of sea ice (Mock et al. 2017). We further found that these new families in *F. cylindrus* have diversified greatly. In detail, CYP5713A1 showed 27% identity to *Danio rerio* CYP7A and the top hits are animal CYP7. This family is pretty ancient, being conserved in animal and choanoflagellates and maybe up to fungi (Nelson et al. 2013). Having this in diatoms too may indicate that this is a much more ancient sequence, with lots of secondary losses (or technical biases like incomplete genome coverage). This is consistent with the basal position of this sequence grouping with CYP5164 and CYP51. CYP5714A1 showed 25% identity to oomycete *Saprolegnia parasitica*. CYP5715A1 showed 20% identity to CYP51C8 o f *C. okamuranus*. CYP5716A1 showed 24% identity to CYP5020A1 of *T. pseudonana*. CYP5717A1 showed 25% identity to CYP5615A10 of *S. parasitica.* CYP5718A1 showed 25% identity to CYP5658A1 of protist *Acanthamoeba castellanii.* CYP5718B1 showed 29% identity to CYP5658A1 of A. *castellanii.* CYP5719A1 showed 22% identity to CYP51C6 of *C. okamuranus*, which also indicates an ancient origin. CYP5720A1 showed 30% identity to CYP5710A1 of *N. gaditana*. The great P450 diversity of this ice diatom might contribute to its cold-adapted character. The maximum length of the CYP proteins occurred in *F. cylindrus* CYP5720A1 of 1273 amino acids (Table S3). P450s fused to redox partners are well documented in the literature (Sello et al. 2015; Guengerich and Munro 2013). The longer sequence of CYP5720A1 came from fusion to P450 reductase in the C-terminal of the protein. This kind of fusion also occurred in the CYP5710A1 of *N. gaditana*. Only twelve CYP protein sequences contained chloroplast-targeting peptides, six of which are CYP97 family members, consistent with the fact that CYP97 are involved in carotenoid metabolism in chloroplast, which plays essential roles in light-harvesting and photoprotection in plant kingdom (Mizutani and Ohta 2010; Tian et al. 2004). . On the other hand, most P450s were found to be endoplasmic reticulum located, which fits the canonical view of P450s bound to the cytoplasmic surface of the endoplasmic reticulum (Mizutani and Ohta 2010; Kelly et al. 2009).

Phylogenetic analysis

To illustrate the relationship among these P450 families, a phylogenetic tree of stramenopile P450s was constructed using the full length of the protein sequences after removing the fused reductase parts. The results showed that some P450s names should be corrected by phylogenetic reconstruction. CYP5015G1, CYP5015G2 and CYP5015G3 of *P. sojae* should be the CYP5015G3, CYP5015G1 and CYP5015G2 orthologs, respectively. CYP5164C3 sequence was changed to CYP5164A4. The CYP5164B and C subfamilies were merged. After these corrections, the distribution on the tree agreed well with their family and subfamily assignment. Genes belonging to the same family grouped together robustly (Fig. 2). Phylogenetic analysis supported the high diversity of P450s in stramenopiles. According to the definition of a clan, where the reproducible clusters end and the variability begins in the deepest branches (Nelson 2017), we defined three clans in stramenopiles, including CYP51 clan shared with plants and animals, CYP97 clan shared with plants, and CYP5014 clan in oomycetes. CYP5715A1 and CYP5713A1 of *F. cylindrus*, CYP5166A1 of *P. tricornutum* and brown algae-specific CYP5164 grouped with the ancient CYP51 clan, reflecting their old origination or slow evolution compared with other families in brown algae or diatoms. CYP5692A2 of *N. gaditana* belonged to CYP5014 clan and showed a close relationship with CYP5017 of oomycetes. CYP5018A1 of *T. pseudonana* weakly grouped with CYP5014 clan. Besides, CYP51C of brown algae formed a larger branch than those of diatoms, resulting from the blooming of CYP51s from *C. okamuranus*. As the most conserved P450 family, CYP51 is widely distributed in all the biological kingdoms and is involved in the essential sterol biosynthesis pathway (Qi et al. 2006). 3 CYPs from brown algae and diatoms should likely come from a direct descendant of the oldest eukaryotic host. Both CYP51 and CYP97 showed a close relationship with red algae, green algae and higher plants. In most organisms that have been examined, only a single CYP51 gene exists, while several plants and fungi contain more than one copy, e.g. rice has ten, *Arabidopsis thaliana* has two, and some *Aspergillus* species have two (Nelson et al. 2004; Lepesheva and Waterman 2007). Most CYP51s across the different kingdoms were thought to have a single strictly conserved function in the synthesis of essential sterols, while some of those additional duplicates (e.g., CYP51H10 in oat) are demonstrated to be involved in the synthesis of antimicrobial compounds avenacins or triterpenes (Qi et al. 2006; Geisler et al. 2013). The identity between the two CYP51s from *A. thaliana* is 72%. It was reported that one CYP51 from *A. thaliana* is an expressed pseudogene, while only CYP51G1 is functional (Kim et al. 2005). Despite the roles of the expanded CYP51s in *C. okamuranus* remain unknown, CYP51C1 of *C. okamuranus* is clearly orthologous to the other two CYP51C1 in *S. japonica* and *E. siliculosus*. This *C. okamuranus* ortholog is 87% identical to CYP51C1 of *S. japonica*, and less than 78% to the other seven *C. okamuranus* CYP51 sequences (Table S4), whereas the identities between the eight CYP51s from *C. okamuranus* range from 51% between CYP51C3 and CYP51C8 to 89% between CYP51C2 and CYP51C4. Furthermore, all the eight CYP51s have the four conserved enzyme functional motifs, suggesting that CYP51s in *C. okamuranus* are probably not pseudogenes but rather are evolving new functions, convergently to what happened in some terrestrial plants.

Despite highly variable sequences of these P450s, the sequence analysis showed some well conserved regions which correspond to the conserved enzyme functional motifs. There are four widely recognized consensus motifs AGXDTT, EXXR, PERF, and FXXGXXXCXG, which are critical for the detection of P450s in genomes (Kelly et al. 2009; Deng et al. 2007). Sequence logo analysis showed that these conserved motifs were present in oomycetes, diatoms and brown algae (Fig. S4). Gene structure analysis revealed that intron numbers vary dramatically between brown algae and diatoms (Fig. S5). The average intron number of

brown algae P450s is seven, whereas that in diatoms is only one. Generally, the genes within the same family have almost the same intron number. In brown algal P450s, the intron number ranged from three in CYP97E8 of *S. japonica* and CYP97E9 of *C. okamuranus* to 13 in CYP5161D1 of *S. japonica*. A large part of brown algal P450 (34%, 17 out of 49) have nine introns. Notably, the CYP51C1 of *S. japonica* and *E. siliculosus*, as well as the seven CYP51C of *C. okamuranus* contain nine introns, indicating the ancestor of brown algae CYP51C contained nine introns. On the contrary, in diatom P450s, the intron number ranged from 0 to 4, with a large part of them (44%, 19 out of 43) have no intron. Csuros et al. globally surveyed 100 complete genomes and inferred that evolution of eukaryotic genes was dominated by intron loss and the ancestor of 'SAR clade' is intron-rich (Csuros et al. 2011). The whole genome of *T. pseudonana* contain an average of 1.4 introns per gene. By evaluation of intron positions conservation, Roy and Penny found *T. pseudonana* has lost the majority of the numerous introns present in the 'SAR clade' ancestor (Roy and Penny 2007). These results revealed that intron loss occurred in diatoms P450s during the course of evolution.

P450 evolutionary dynamics in brown algae and diatoms

Based on naming and classification of families and subfamilies, further dynamics study was performed to assess the P450 evolutionary history within photosynthetic stramenopiles ochrophytes. Using the Dollo parsimony principle (Farris 1977; Almeida et al. 2016), combined with the results from 1KP transcriptome database, we estimated the P450 families that were gained or lost at each node of the algae tree(Fig. 3). In the common ancestor of brown algae and diatoms, there were six P450 families, including CYP51, CYP97, CYP5160, CYP5021, CYP5022, and CYP5165. Then, three families, that is, CYP5021, CYP5022, and CYP5165 were lost in brown algae. CYP5161, CYP5162, and CYP5164 were acquired in the ancestor of brown algae, followed by acquisition of CYP5163 in the ancestor of *E. siliculosus* and *C. okamuranus*. Alternatively, some of those CYPs may be highly diverging forms of otherwise conserved family. This is at least the case for CYP5164B1 that branches with CYP74 from eukaryotes and bacteria (Toporkova et al. 2017). Diatoms can produce a various set of epoxy alcohols, similar to those made by brown algae (Andreou et al. 2009). Thus, if they have no CYP74/CYP5164, this mean that either this enzyme diverged in terms of sequence until the point that it is not recognizable, or that allene oxide synthase activity was reacquired in diatoms from a different CYP paralog. The same is true for oomycetes, which are also able to produce epoxy alcohols from PUFAs (Andreou et al. 2009). The diatoms, on the other hand, seem to have undergone more frequent gain or loss events. CYP5023 was acquired in their ancestor. Subsequently, *T. pseudonana* gained four families, CYP5018, CYP5019, CYP5020 and CYP5024, and lost two families CYP5160 and CYP5165. Then, *P. tricornutum* gained CYP5166 and lost CYP5022 and CYP5023, whereas *F. cylindrus* lost CYP5160 and gained families CYP5712-5720. At last, diatom *F. cylindrus* contains the most CYP families of 15, followed by nine families in *T. pseudonana*, seven families in *E. siliculosus* and *C. okamuranus*, six families in *S. japonica* and *P. tricornutum*. Though drawing this evolutionary process, we cannot rule out other possibilities due to the biases linked to birth/death estimations in the context of multiple losses and heterogeneity in genome coverage. The dynamic changes of P450 families revealed that brown algae and diatoms evolved distinct P450 composition in their genomes, strongly suggesting their evolutionary adaptation to different coastal ecosystem.

6lyOn the other hand, tandem duplication rarely occurred in diatoms examined, with only CYP5165A1 and CYP5165A2 on chromosome 1 of *P. tricornutum*.

Higher evolutionary rate may reflect the adaption to different substrates or emerging new functions. Lower evolutionary rate, on the other hand, would indicate the functional conservation or a strict substrate specificity (Parvez et al. 2016). To gain some insight into evolutionary rate and selection pressure on the evolution of the P450s, we compared ω values (ratio of non-synonymous to synonymous codon substitutions, d*N*/d*S*) in main families of brown algal and diatom P450. The one-ratio model of Codeml (one ω ratio for each family tree) showed that all the P450 families examined were have an ω of less than 0.4 (Fig. 4a). Under the site model that allows the ω ratio to vary among sites (M7 and M8), no positive selected site was detected (Table S5). CYP5163 sequences of brown algae had higher ω value than other families, maybe in relation with the fact it emerged recently , given that ω values decay over time, due to the accumulation of synonymous substitutions. CYP51 and CYP97 in diatoms have lower ω values than those of brown algae, indicating the slow evolutionary rates in diatoms. Considering that the higher ω ratio in brown algal CYP51 may result from the expansion of CYP51 in *C. okamuranus*, we further applied the free-ratio model (independent ω ratio for each branch) to test the selection along each branch (Fig. 4b). Interestingly, we detected higher ω value along the branches leading to the *C. okamuranus* CYP51C2-C8 compared with other CYP51C1 genes. To further confirm this result, branch specific model (M = 2) and branch-site model (M = 2, Nssites = 2) were used to check the selection pressure of the expanded CYP51Cs in *C. okamuranus*. Consistently, the CYP51C2-C8 clade exhibited higher ω value (ω = 0.5180) than other clades, though no site under positive selection was detected (Table S5). This result suggested that the relaxed selection pressure might contribute to the remaining of multiple CYP51C in *C. okamuranus* and further to acquisition of new activity or a change of function emerging post-duplication. On the contrary, the CYP51s of diatoms and *N. gaditana* were found under strong negative selection, indicating a functional conservation in the course of evolution.

Transcriptome analysis and evolution of two critical steps in the oxylipin synthesis pathway among stramenopiles

We used publicly available microarray datasets from *E. siliculosus* (Dittami et al. 2009; Ritter et al. 2014), diatoms (Ashworth et al. 2016) and our RNA-seq datasets of the three life stages of *S. japonica* including sporophytes, male gametophytes and female gametophytes to investigate the expression profiles of *P450* genes. Among the fifteen *P450*s of *S. japonica*, four genes were not expressed in all the three life stages, two of which are truncated pseudogenes (Fig. 5a). The other two (*CYP5161C3* and *CYP5164B3*) may have become pseudogenes after duplication or evolved new function not identified in our study, such as expression induced by herbivore grazing, as it may occur with other kelps (Ritter et al. 2017). Seven genes showed significantly higher expression levels in sporophytes (adjusted p value < 0.05), including *CYP5164B2*, *CYP5161C1*, *CYP5161C2*, *CYP5164C1*, *CYP51C1*, *CYP5161D1* and *CYP5160B2*. In previous study, Toporkova et al. (2017) cloned *CYP5164B1* of *E. siliculosus* and showed it worked as epoxyalcohol synthase, converting fatty acid hydroperoxides into oxylipin epoxyalcohols. Brown algae possess a great diversity of oxylipins (Ritter et al. 2008; Proteau and Gerwick 1992). Considering that *CYP5164*s expanded in *S. japonica* and is mainly expressed in sporophytes, they may contribute to high amounts of oxylipins biosynthesis in this large kelp. Besides, only one gene, *CYP97F8* was highly expressed in male and female gametophytes, suggesting its pivotal roles in gametophyte development.

Transcriptome analysis of *E. siliculosus* showed that most *P450* genes are responsive to at

least one stress condition (fold change > 2 compared with control) (Fig. 5b). Ritter et al. demonstrated that copper treatment (Cu) could induce oxidative stress and activate oxylipin biosynthesis in *E. siliculosus* and kelp *Laminaria digitata* (Ritter et al. 2008; Ritter et al. 2014). Two genes are responsive to Cu stress, with one gene *CYP5164A2* up-regulated while the other gene *CYP97F4* down-regulated, suggesting that *CYP5164A2* either participates in oxylipin synthesis, or in detoxification of environmental chemicals. Moreover, *CYP5164A2* was also up-regulated under oxidative and hyposaline stresses, which further supports the anti-oxidation roles of this gene. Besides, *CYP5164B1* of *E. siliculosus*, which has been proved to synthesize oxylipin epoxyalcohol (Toporkova et al. 2017) was up-regulated under hypersaline stress while down-regulated under oxidative stress, indicating it is involved in stress response. Additionally, *CYP97F4*, which participates in photosynthetic pigment synthesis, was down-regulated under all conditions, indicating that photosynthesis was vulnerable under these stress conditions.

Diatoms transcriptomic analysis revealed some CYPs are responsive to specific conditions. Similar to *E. siliculosus*, *CYP97* in diatoms were also down regulated across many conditions, such as silicon, iron and nitrogen limitation in *T. pseudonana*, short term high light and blue light treatment in *P. tricornutum* (Fig. 5c,d), probably reflecting a negative-feedback regulation of photosynthesis. Besides, in *T. pseudonana*, *CYP5024A1* and *CYP5018A1* were up regulated under low temperature and down regulated under silicon limitation, *CYP5019A1* was up regulated when exposed to the pollutant benzo[a]pyrene at a sub-lethal concentration. On the other hand, in *P. tricornutum*, strong and rapid response to the changes in irradiance was observed in *CYP5166A1* and *CYP5165A1*. *CYP5166A1* was up regulated after transfer to high light conditions, indicating its potential roles in protecting the organism from photo-oxidative damage to the photosynthetic machinery. *CYP5165A1* was up regulated under red, green and blue light at the initial response phase, followed by the intermediate and late acclimation phase, suggesting its critical roles in sensing spectral quality.

Our new transcriptomic data on P450 genes, as well as our previous studies on the LOX gene family (Teng et al., 2017b) can now be integrated with metabolic data to propose a first comparative model of the first two steps in the oxylipin synthesis pathway among stramenopiles (Fig. 6). Lineage-specific models have already been published previously, at a time when few whole-genome data were available for stramenopiles (Andreou et al., 2009). Now both genomic and metabolomic data are available for *Ectocarpus siliculosus*, and genome and metabolome data are available for two species pairs from the same genus in two other stramenopiles. Genome data from *S. japonica* can be compared to metabolome data from *S. angustata* (Boonprab et al. 2003b; Boonprab et al. 2003a), whereas genome data from *T. pseudonana* can be related with metabolic data from *T. rotula* (Barofsky and Pohnert 2007; D'Ippolito et al. 2006). It is not yet possible to assign all P450 paralogs to a specific enzymatic activity, but our data already provide some clarifications. First, absence of a CYP74/CYP5164-like gene in Thalassosiria correlates with absence of an hydroperoxide lyase (HPL) activity linked to a P450. On the contrary, CYP5713A1 from another diatom, *Fragillariopsis cylindricus*, would be an interesting candidate to test if a P450-linked HPL activity is still retained in some diatoms. Second, it is already clear that land plants and brown algae independently underwent diversification of either CYP74 or CYP5164. From that we can expect that the biochemical activities are differently subfunctionalized across the angiosperm and stramenopile lineages.

Conclusions

In this study, we systematically investigated the P450s diversity and evolution within stramenopiles species, which evolved from secondary endosymbiosis events. Three clans were defined in this lineage, including CYP51 clan, CYP97 clan and CYP5014 clan. CYPomes characteristics differ greatly among oomycetes, brown algae and diatoms. Phylogenetic analysis revealed that P450s of diatoms and brown algae showed a closer relationship than with oomycetes P450s. The CYP51 family expanded in brown alga *C. okamuranus*. 1KP transcriptome analysis revealed that some families, such as CYP5021 and CYP5022 are not restricted to diatoms. Gene gain and loss patterns revealed that six CYP families – CYP51, CYP97 and CYP5160, CYP5021, CYP5022 and CYP5165 predated the split of brown algae and diatoms. After they diverged, diatoms gained more CYP families, especially for the cold-adapted species *F. cylindrus*, in which eight new CYP families were found. Expression analysis showed that most P450s in *S. japonica* were highly expressed in sporophytes, which suggests a potential role in large kelp development. *E. siliculosus* and diatom P450s, on the other hand, were found responsive to many stressful or irradiance conditions, suggesting their important roles in response to environmental stimulus. The information provided in this study will be helpful in further investigating P450 roles in this diversified lineage.

Acknowledgements

This work was supported by Central Public-interest Scientific Institution Basal Research Fund CAFS (2017HY-YJ01); National key Research and Development Plan (2016YFC1402102-2); National Natural Science Foundation of China (41676145); Shandong key Research and Development Plan; Qingdao Municipal Science and Technology plan project (17-1-1-96-jch); Special Scientific Research Funds for Central Non-Profit Institutes, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences (20603022016010, 20603022016001); China Agriculture Research System (CARS-50); Taishan Scholars Funding; AoShan Talents Program (No. 2015ASTPES03); the Science Fund for Distinguished Young Scholars of Shandong Province (JQ201509); and Projects of International Exchange and Cooperation in Agriculture, Ministry of Agriculture and Rural Affairs of China-Science, Technology and Innovation Cooperation in Aquaculture with Tropical Countries along the Belt and Road. G.V.M. benefited from the support of the French Government via the National Research Agency investment expenditure program IDEALG (ANR-10-BTBR-04) and from Région Bretagne via the grant « SAD2016-METALG (9673) ».

Reference

Aleoshin VV, Mylnikov AP, Mirzaeva GS, Mikhailov KV, Karpov SA (2016) Heterokont PredatorDevelorapax marinusgen. et sp. nov. – A Model of the Ochrophyte Ancestor. Frontiers in Microbiology 7 (16)

Almeida D, Maldonado E, Khan I, Silva L, Gilbert MTP, Zhang G, Jarvis ED, O'Brien SJ, Johnson WE, Antunes A (2016) Whole-Genome Identification, Phylogeny, and

Evolution of the Cytochrome P450 Family 2 (CYP2) Subfamilies in Birds. Genome Biology & Evolution 8 (4):1115-1131
Andreou A, Brodhun F, Feussner I (2009) Biosynthesis of oxylipins in non-mammals. Progress in Lipid Research 48 (3–4):148-170
Ashworth J, Turkarslan S, Harris M, Orellana MV, Baliga NS (2016) Pan-transcriptomic analysis identifies coordinated and orthologous functional modules in the diatoms Thalassiosira pseudonana and Phaeodactylum tricornutum. Marine Genomics 26:21
Barofsky A, Pohnert G (2007) Biosynthesis of polyunsaturated short chain aldehydes in the diatom Thalassiosira rotula. Organic Letters 9 (6):1017-1020
Benjamini Y, Hochberg Y (1995) Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. Journal of the Royal Statistical Society 57 (1):289-300
Boonprab K, Matsui K, Akakabe Y, Yotsukura N, Kajiwara T (2003a) Hydroperoxy-arachidonic acid mediated n-hexanal and (Z)-3- and (E)-2-nonenal formation in Laminaria angustata. Phytochemistry 63 (6):669
Boonprab K, Matsui K, Yoshida M, Akakabe Y, Chirapart A, Kajiwara T (2003b) C6-aldehyde formation by fatty acid hydroperoxide lyase in the brown alga Laminaria angustata. Zeitschrift Für Naturforschung C 58 (4):207-214
Brembu T, Jørstad M, Winge P, Valle KC, Bones AM (2011) Genome-Wide Profiling of Responses to Cadmium in the Diatom Phaeodactylum tricornutum. Environmental Science & Technology 45 (18):7640-7647
Carvalho RN, Bopp SK, Lettieri T (2011) Transcriptomics Responses in Marine Diatom Thalassiosira pseudonana Exposed to the Polycyclic Aromatic Hydrocarbon Benzo[a]pyrene. Plos One 6 (11):: e26985.
Chou K-C, Shen H-B (2010) A new method for predicting the subcellular localization of eukaryotic proteins with both single and multiple sites: Euk-mPLoc 2.0. PLoS One 5 (4):e9931
Cock JM, Sterck L, Ahmed S, Allen A, Amoutzias G, Anthouard V, Artiguenave F, Arun A, Aury JM (2012) The Ectocarpus genome and brown algal genomics. Elsevier Academic Press, San Diego, CA
Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, Anthouard V, Artiguenave F, Aury JM, Badger JH (2010) The Ectocarpus genome and the independent evolution of multicellularity in brown algae. Nature 465:617-621
Cormier A, Avia K, Sterck L, Derrien T, Wucher V, Andres G, Monsoor M, Godfroy O, Lipinska A, Perrineau M-M, Peer YVD, Hitte C, Corre E, Coelho SM, Cock JM (2016) Re-annotation, improved large-scale assembly and establishment of a catalogue of noncoding loci for the genome of the model brown alga Ectocarpus. New Phytologist 214 (1)
Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A Sequence Logo Generator. Genome Research 14 (6):1188-1190
Csűös M (2010) Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. Bioinformatics 26 (15):1910-1912
Csuros M, Rogozin IB, Koonin EV (2011) A detailed history of intron-rich eukaryotic ancestors inferred from a global survey of 100 complete genomes. PLoS computational biology 7 (9):e1002150
D'Ippolito G, Cutignano A, Tucci S, Romano G, Cimino G, Fontana A (2006) Biosynthetic intermediates and stereochemical aspects of aldehyde biosynthesis in the marine diatom Thalassiosira rotula. Phytochemistry 67 (3):314-322
Debeljak N, Horvat S, Vouk K, Lee M, Rozman D (2000) Characterization of the mouse lanosterol 14α-demethylase (CYP51), a new member of the evolutionarily most

conserved cytochrome P450 family. Archives of biochemistry and biophysics 379 (1):37-45
Deng J, Carbone I, Dean RA (2007) The evolutionary history of Cytochrome P450 genes in four filamentous Ascomycetes. Bmc Evolutionary Biology 7 (1):30
Dittami SM, Scornet D, Petit J-L, Ségurens B, Da Silva C, Corre E, Dondrup M, Glatting K-H, König R, Sterck L (2009) Global expression analysis of the brown alga Ectocarpus siliculosus (Phaeophyceae) reveals large-scale reprogramming of the transcriptome in response to abiotic stress. Genome biology 10 (6):R66
Dorrell RG, Gile G, Mccallum G, Méheust R, Bapteste EP, Klinger CM, Brillet-Guéguen L, Freeman KD, Richter DJ, Bowler C (2017) Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. eLife 6:e23717
Eddy SR (2011) Accelerated Profile HMM Searches. Plos Computational Biology 7 (10):e1002195
Farris JS (1977) Phylogenetic Analysis Under Dollo's Law. Systematic Zoology 26 (1):77-88
Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A (2016) The Pfam protein families database: towards a more sustainable future. Nucleic acids research 44 (D1):D279-D285
Frommolt R, Werner S, Paulsen H, Goss R, Wilhelm C, Zauner S, Maier UG, Grossman AR, Bhattacharya D, Lohr M (2008) Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis. Molecular Biology & Evolution 25 (12):2653-2667(2615)
Geisler K, Hughes RK, Sainsbury F, Lomonossoff GP, Rejzek M, Fairhurst S, Olsen CE, Motawia MS, Melton RE, Hemmings AM (2013) Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. Proceedings of the National Academy of Sciences of the United States of America 110 (35):E3360
Goldstone HM, Stegeman JJ (2006) A revised evolutionary history of the CYP1A subfamily: gene duplication, gene conversion, and positive selection. Journal of Molecular Evolution 62 (6):708-717
Grigoriev IV, Nordberg H, Shabalov I, Aerts A, Cantor M, Goodstein D, Kuo A, Minovitsky S, Nikitin R, Ohm RA (2014) The Genome Portal of the Department of Energy Joint Genome Institute. Nucleic Acids Research 42 (Database issue):26-31
Gruber A, Rocap G, Kroth PG, Armbrust E, Mock T (2015) Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. The Plant Journal 81 (3):519-528
Guengerich FP, Munro AW (2013) Unusual cytochrome p450 enzymes and reactions. Journal of Biological Chemistry 288 (24):17065-17073
Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G (2014) GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 31 (8):1296-1297
Keeling PJ (2013) The Number, Speed, and Impact of Plastid Endosymbioses in Eukaryotic Evolution. Annual Review of Plant Biology 64 (1):583-607
Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV, Archibald JM, Bharti AK, Bell CJ (2014) The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. Plos Biology 12 (6):429-440
Kelly DE, Kraševec N, Mullins J, Nelson DR (2009) The CYPome (cytochrome P450 complement) of Aspergillus nidulans. Fungal genetics and biology 46 (1):S53-S61
Kersey PJ, Allen JE, Allot A, Barba M, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Grabmueller C (2018) Ensembl Genomes 2018: an integrated omics

infrastructure for non-vertebrate species. Nucleic Acids Research 46 (Database issue):D802-D808
Kim HB, Schaller H, Goh C-H, Kwon M, Choe S, An CS, Durst F, Feldmann KA, Feyereisen R (2005) Arabidopsis cyp51 mutant shows postembryonic seedling lethality associated with lack of membrane integrity. Plant physiology 138 (4):2033-2047
Kousaka K, Ogi N, Akazawa Y, Fujieda M, Yamamoto Y, Takada Y, Kimura J (2003) Novel oxylipin metabolites from the brown alga Eisenia bicyclis. Journal of natural products 66 (10):1318-1323
Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular biology and evolution 33 (7):1870-1874
Lepesheva GI, Waterman MR (2007) Sterol 14α-demethylase cytochrome P450 (CYP51), a P450 in all biological kingdoms. Biochimica et Biophysica Acta (BBA)-General Subjects 1770 (3):467-477
Li Z, Defoort J, Tasdighian S, Maere S, Van de Peer Y, De Smet R (2016) Gene Duplicability of Core Genes Is Highly Consistent across All Angiosperms. Plant Cell 28 (2):326-344. doi:10.1105/tpc.15.00877
Liu Z, Tavares R, Forsythe ES, André F, Lugan R, Jonasson G, Boutet-Mercey S, Tohge T, Beilstein MA, Werck-Reichhart D, Renault H (2016) Evolutionary interplay between sister cytochrome P450 genes shapes plasticity in plant metabolism. Nature Communications 7:13026
Lobo I (2012) Basic Local Alignment Search Tool (BLAST). Journal of Molecular Biology 215 (3):403-410
Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15 (12):550
Mizutani M, Ohta D (2010) Diversification of P450 Genes During Land Plant Evolution. Annual Review of Plant Biology 61 (1):291
Mock T, Otillar RP, Strauss J, McMullan M, Paajanen P, Schmutz J, Salamov A, Sanges R, Toseland A, Ward BJ (2017) Evolutionary genomics of the cold-adapted diatom Fragilariopsis cylindrus. Nature:1-5
Mock T, Samanta MP, Iverson V, Berthiaume C, Robison M, Holtermann K, Durkin C, Bondurant SS, Richmond K, Rodesch M (2008) Whole-Genome Expression Profiling of the Marine Diatom Thalassiosira pseudonana Identifies Genes Involved in Silicon Bioprocesses. Proceedings of the National Academy of Sciences of the United States of America 105 (5):1579-1584
Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D (2009) Genomic Footprints of a Cryptic Plastid Endosymbiosis in Diatoms. Science 324. doi:10.1126/science.1172983
Nelson DR (1998) Cytochrome P450 nomenclature. Methods in Molecular Biology 107:15
Nelson DR (2009) The Cytochrome P450 Homepage. Human Genomics 4 (1):59
Nelson DR (2017) Cytochrome P450 diversity in the tree of life. Biochimica Et Biophysica Acta
Nelson DR, Goldstone JV, Stegeman JJ (2013) The cytochrome P450 genesis locus: the origin and evolution of animal cytochrome P450s. Philosophical Transactions of the Royal Society of London 368 (1612):20120474
Nelson DR, Schuler MA, Paquette SM, Werck-Reichhart D, Bak S (2004) Comparative genomics of rice and Arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. Plant Physiology 135 (2):756-772
Nishitsuji K, Arimoto A, Iwai K, Sudo Y, Hisata K, Fujie M, Arakaki N, Kushiro T, Konishi T, Shinzato C (2016) A draft genome of the brown alga, Cladosiphon okamuranus, S-strain: a platform for future studies of ‘mozuku’biology. Dna Research 23 (6):561-570

Nymark M, Valle KC, Brembu T, Hancke K, Winge P, Andresen K, Johnsen G, Bones AM (2009) An Integrated Analysis of Molecular Acclimation to High Light in the Marine Diatom Phaeodactylum tricornutum. Plos One 4 (11):e7743

Omura T, Gotoh O (2017) Evolutionary Origin of Mitochondrial Cytochrome P450. Journal of Biochemistry 161 (5):399-407

Parvez M, Qhanya LB, Mthakathi NT, Kgosiemang IK, Bamal HD, Pagadala NS, Xie T, Yang H, Chen H, Theron CW (2016) Molecular evolutionary dynamics of cytochrome P450 monooxygenases across kingdoms: Special focus on mycobacterial P450s. Scientific Reports 6 (33099):33099

Pflugmacher S, Sandermann H (1998) Cytochrome P450 monooxygenases for fatty acids and xenobiotics in marine macroalgae. Plant physiology 117 (1):123-128

Phillips IR, Shephard EA (1998) Cytochrome P450 protocols. Humana Press, New Jersey

Proteau PJ, Gerwick WH (1992) Cymathere ethers A and B: bicyclic oxylipins from the marine brown alga Cymathere triplicata. Tetrahedron letters 33 (31):4393-4396

Qi X, Bakht S, Qin B, Leggett M, Hemmings A, Mellon F, Eagles J, Werck-Reichhart D, Schaller H, Lesot A (2006) A different function for a member of an ancient and highly conserved cytochrome P450 family: From essential sterols to plant defense. Proceedings of the National Academy of Sciences of the United States of America 103 (49):18848-18853

Renault H, Bassard JE, Hamberger B, Werck-Reichhart D (2014) Cytochrome P450-mediated metabolic engineering: current progress and future challenges. Current Opinion in Plant Biology 19 (19C):27-34

Renault H, De MM, Jonasson G, Lara P, Nelson DR, Nilsson I, André F, Von HG, Werck-Reichhart D (2017) Gene duplication leads to altered membrane topology of a cytochrome P450 enzyme in seed plants. Molecular Biology & Evolution

Ritter A, Cabioch L, Brilletguéguen L, Corre E, Cosse A, Dartevelle L, Duruflé H, Fasshauer C, Goulitquer S, Thomas F (2017) Herbivore-induced chemical and molecular responses of the kelps Laminaria digitata and Lessonia spicata. Plos One 12 (3):e0173315

Ritter A, Dittami SM, Goulitquer S, Correa JA, Boyen C, Potin P, Tonon T (2014) Transcriptomic and metabolomic analysis of copper stress acclimation in Ectocarpus siliculosus highlights signaling and tolerance mechanisms in brown algae. BMC plant biology 14 (1):116

Ritter A, Goulitquer S, Salaün JP, Tonon T, Correa JA, Potin P (2008) Copper stress induces biosynthesis of octadecanoid and eicosanoid oxygenated derivatives in the brown algal kelp Laminaria digitata. New Phytologist 180 (4):809-821

Roy SW, Penny D (2007) A very high fraction of unique intron positions in the intron-rich diatom Thalassiosira pseudonana indicates widespread intron gain. Molecular biology and evolution 24 (7):1447-1457

Sapriel G, Quinet M, Heijde M, Jourdren L, Tanty V, Luo G, Le Crom S, Lopez P (2009) Genome-wide transcriptome analyses of silicon metabolism in Phaeodactylum tricornutum reveal the multilevel regulation of silicic acid transporters. PLoS ONE 4 (10):e7458

Sello MM, Jafta N, Nelson DR, Chen W, Yu JH, Parvez M, Kgosiemang IKR, Monyaki R, Raselemane SC, Qhanya LB (2015) Diversity and evolution of cytochrome P450 monooxygenases in Oomycetes. Scientific Reports 5: 11572

Ševčíková T, Horák A, Klimeš V, Zbránková V, Demir-Hilton E, Sudek S, Jenkins J, Schmutz J, Přibyl P, Fousek J, Vl

ek Č, Lang BF, Oborník M, Worden AZ, Eliá š M (2015) Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge

through endosymbiosis of an ochrophyte? Scientific Reports 5:10134. doi:10.1038/srep10134
Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J (2011) Fast, scalable generation of high- quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology 7 (1):539
Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic acids research 34 (suppl_2):W609-W612
Syed K, Nelson DR, Riley R, Yadav JS (2013) Genomewide annotation and comparative genomics of cytochrome P450 monooxygenases (P450s) in the polypore species Bjerkandera adusta, Ganoderma sp. and Phlebia brevispora. Mycologia 105 (6):1445-1455. doi:10.3852/13-002
Teng L, Fan X, Xu D, Zhang X, Mock T, Ye N (2017a) Identification of Genes under Positive Selection Reveals Differences in Evolutionary Adaptation between Brown-Algal Species. Frontiers in Plant Science 8:1429
Teng L, Han W, Fan X, Xu D, Zhang X, Dittami SM, Ye N (2017b) Evolution and Expansion of the Prokaryote-Like Lipoxygenase Family in the Brown Alga Saccharina japonica. Frontiers in Plant Science 8:2018
Tian L, Musetti V, Kim J, Magallaneslundback M, Dellapenna D (2004) The Arabidopsis LUT1 locus encodes a member of the cytochrome P450 family that is required for carotenoid ε-ring hydroxylation activity. Proceedings of the National Academy of Sciences of the United States of America 101 (1):402
Toporkova YY, Fatykhova VS, Gogolev YV, Khairutdinov BI, Mukhtarova LS, Grechkin AN (2017) Epoxyalcohol synthase of Ectocarpus siliculosus. First CYP74-related enzyme of oxylipin biosynthesis in brown algae. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids 1862 (2):167-175
Valle KC, Nymark M, Aamot I, Hancke K, Winge P, Andresen K, Johnsen G, Brembu T, Bones AM (2014) System Responses to Equal Doses of Photosynthetically Usable Radiation of Blue, Green, and Red Light in the Marine Diatom Phaeodactylum tricornutum. Plos One 9 (12):e114211
Voorrips RE (2002) MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. Journal of Heredity 93 (1):77-78
Yang S, Wu RS, Mok HO, Zhang Z, Kong RY (2003) Identification of a novel cytochrome P450 cDNA, CYP97E1, from the marine diatom Skeletonema costatum Bacillariophyceae. Journal of Phycology 39 (3):555-560
Yang Z (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 24 (8):1586-1591
Ye N, Zhang X, Miao M, Fan X, Zheng Y, Xu D, Wang J, Zhou L, Wang D, Gao Y (2015) Saccharina genomes provide novel insight into kelp biology. Nature communications 6:6986
Yu L, Tang W, He W, Ma X, Vasseur L, Baxter SW, Yang G, Huang S, Song F, You M (2015) Characterization and expression of the cytochrome P450 gene family in diamondback moth, Plutella xylostella (L.). Scientific Reports 5:8952
Zhao H, Ma H, Yu L, Wang X, Zhao J (2012) Genome-Wide Survey and Expression Analysis of Amino Acid Transporter Gene Family in Rice (Oryza sativa L.). Plos One 7 (11):e49210

Figure legend

Fig. 1 Statistics of P450 families in each stramenopile species. **a** CYP family member distribution. **b** Comparative analysis of the number of P450s and P450 families. C. P450 diversity percentage. It is represented by the percentage of the number of P450 families in the total number of P450s in an organism.

Fig. 2 Phylogenetic tree of P450s in stramenopile species. The ML tree was constructed using MEGA 7.0 with LG+G+F model. The branch names represent the P450 name, followed by the species names. The branches are colored to show their different taxonomic groups of oomycetes, brown algae, diatoms and *N. gaditana*, respectively. The number on the tree node means bootstrap (1000 replicates) values above 50%.

Fig. 3 Evolutionary process of brown algae/diatoms CYP families. Green circles indicate the gain of the corresponding CYP families while red circles indicate their loss. Numbers along the phylogenetic tree denote the number of CYP families present in each moment of the alga evolution

Fig. 4 ω values computed with Codeml branch model. **a** The ω values of CYP families using branch one ratio model (M=0). **b** The ω values in the CYP51 family. Blue number is the ω values of each gene under free ratio model (M=1), red number is the ω values of different phylogenetic groups under specific branch model (M=2).

Fig. 5 Expression profiles of P450 genes. **a** Log10-transformed FPKM values in *S. japonica*. Black stars indicate sex-biased or sporophyte-biased expression (fold changes > 2, adjusted *p*-value < 0.05). SP: sporophytes; FG: female gametophytes; MG: male gametophytes; **b** Log2-transformed fold changes of expression levels compared to the control in *E. siliculosus*. hyper: hypersaline; hypo: hyposaline; oxi: oxidative; Cu_4: copper treated for 4 h; Cu_8: copper treated for 8 h. Black stars indicate the log2 values of fold changes (treatments/control) > 1.0 or <-1.0; **c** Log2-transformed fold changes of expression levels compared to the control in *T. pseudonana.* T, Si, Fe, N: low temperature, silicon, iron and nitrogen limitation; BaP: benzo[a]pyrene; pH: alkaline pH; Black stars indicate the log2 values of fold changes (treatments/control) > 1.0 or <-1.0; **d** Log2-transformed fold changes of expression levels compared to the control in *P. tricornutum*. Black stars indicate the log2 values of fold changes (treatments/control) > 1.0 or <-1.0. 'NA' means the data is not available.

Fig. 6 Comparison of the two steps in the oxilipin synthesis pathway across stramenopiles. Each of the color frame represents a synthetic pathway in one lineage. This is a single-species lineage for *A. thaliana* and *E. siliculosus*, whereas genomic and metabolomic data are pooled for two species from the same genus for *Saccharina* and *Thalassiosiria*. The peroxidation step in orange is always carried out by a LOX enzyme, wherease the hydroperoxilyase activity (in purple) can be carried out either by a P450 (*Arabidopsis* and the two brown algae) or by an enzyme from a different family in *Thalassiosiria* (blue frame).

Supplementary material

Fig. S1 Statistics of P450 families in each stramenopiles species.

Fig. S2 Phylogenetic tree of CYPs in brown algae, diatoms, *Bolidomonas* and *Vaucheria*. The ML tree was constructed using MEGA 7.0 with WAG+G model. The dark yellow branches are CYPs only

occurred in brown algae. The blue branches are CYPs only occurred in diatoms. The red and green branched correspond to CYPs from *Bolidomonas* and *Vaucheria*, respectively.

Fig. S3 Phylogeny of brown algae and diatoms CYP51 and CYP97 with their counterparts in animal, plant, bacterial and fungi. The phylogenetic tree was constructed with MEGA 7 and displayed with Figtree, using the alignments of CYPs generated by hmmalign in HMMER program with CYP sequences and PF00067.hmm as input files. 16 animal CYP51s, 10 bacteria CYP51s, 27 plant CYP51s, 21 fungal CYP51s and 16 plant CYP97s were extracted from the CYP homepage http://drnelson.uthsc.edu/CytochromeP450.html.

Fig. S4 Sequence logos of the conserved CYP motifs from the three lineages of oomycetes, diatoms and brown algae. Multiple alignments of CYP proteins were performed by aligning them to the PF00067.hmm with HMMER package. The consensus logos were generated by Weblogo http://weblogo.threeplusone.com/create.cgi.

Fig. S5 The gene structure of P450 genes. **a** The gene structure of P450 genes from brown algae. **b** The gene structure of P450 genes from diatoms. **c** The boxplot showing the intron count distribution.

Fig. S6 Chromosome or scaffold distribution of P450s in brown algae and diatom. The position of each P450 was mapped to the genome. Es: *E. siliculosus*, Sj: *S. japonica*, Co: *C. okamuranus*, Pt: *Phaeodactylum tricornutum*, Tp: *Thalassiosira pseudonana*

Table S1 Statistics of P450 number in each species.

Table S2 Results of BLAST against 1KP transcriptome database using CYPs only occurring in brown algae (CYP5161, CYP5162, CYP5163 and CYP5164) and CYPs only occurring in diatoms (CYP5021, CYP5022 and CYP5165)

Table S3 CYP list, showing the CYP name, protein length, subcellular location, and transmembrane helix number (TH).

Table S4 Sequence identity between brown algal CYP51.

Table S5 Statistics of Codeml calculation.

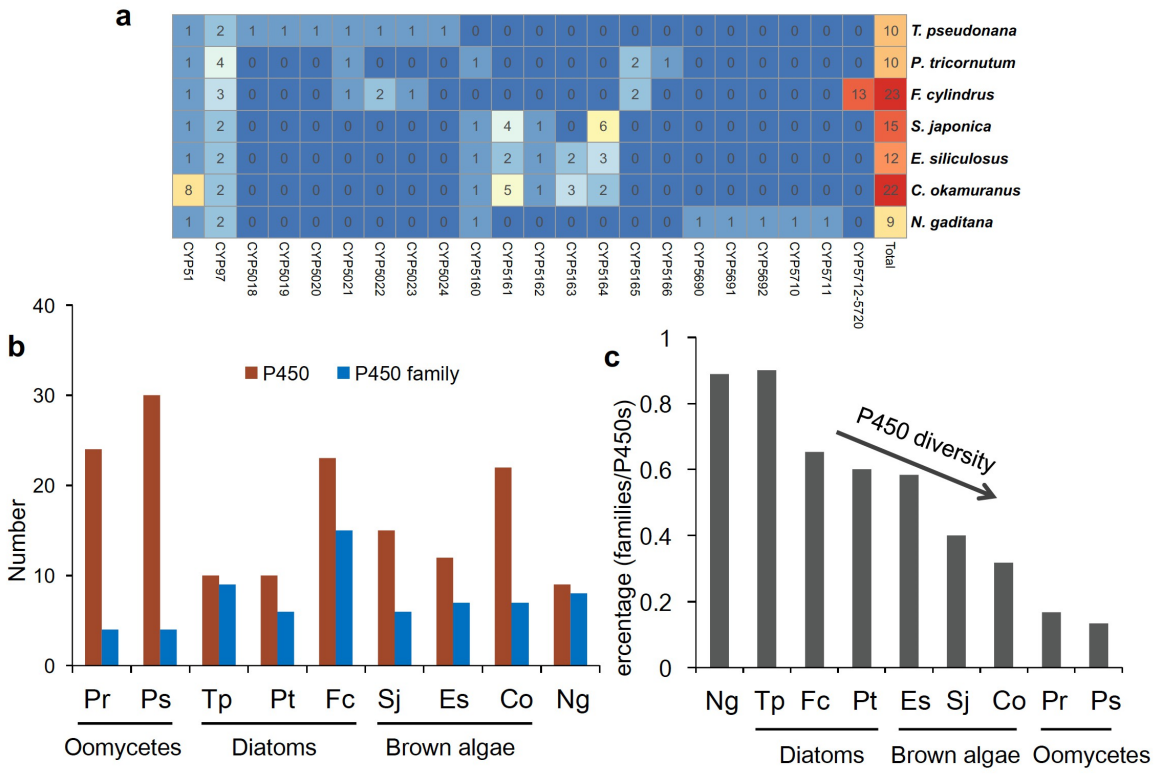


fig.1

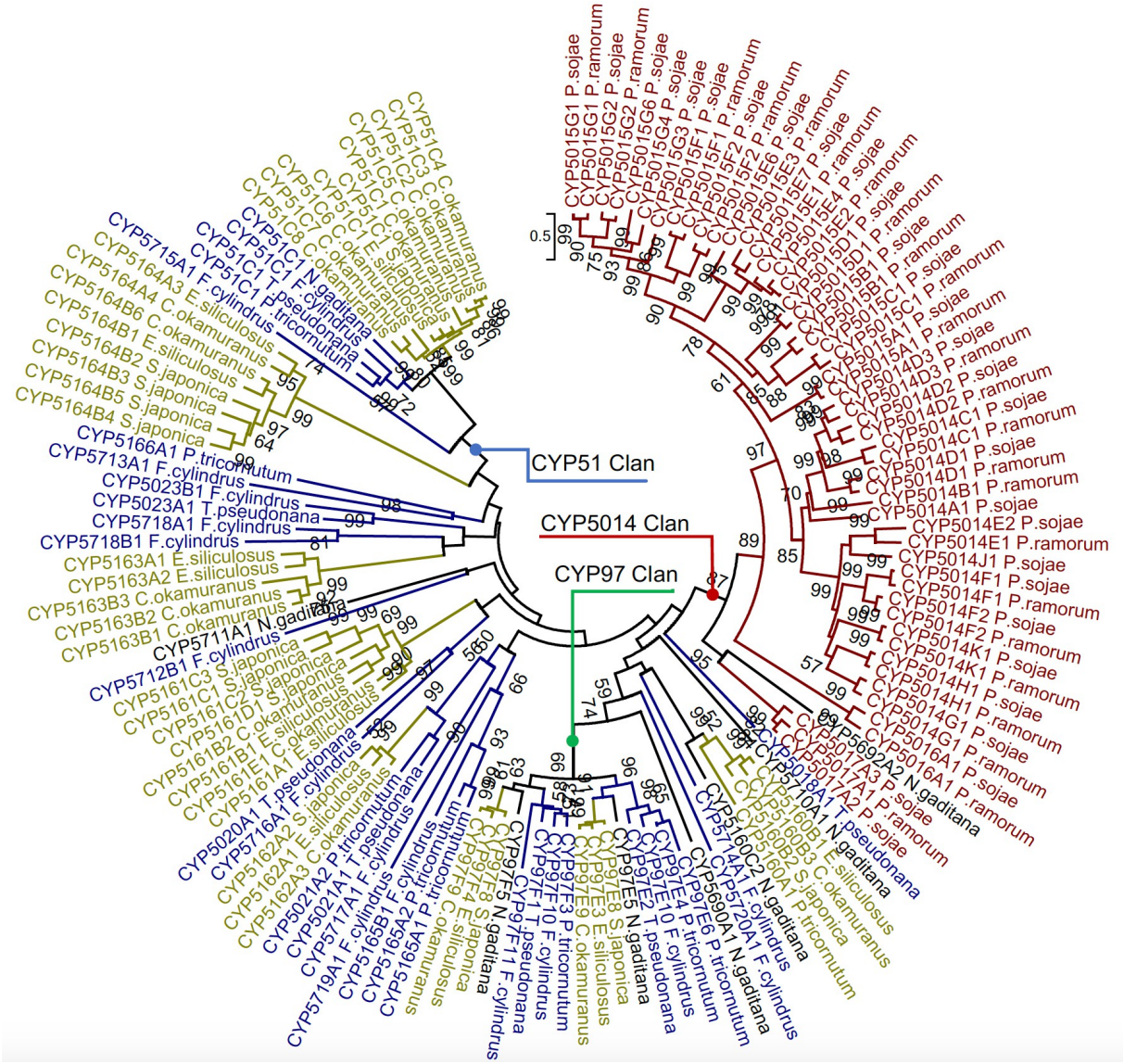

CYP51 Clan
CYP5014 Clan
CYP97 Clan
0.5

fig.2

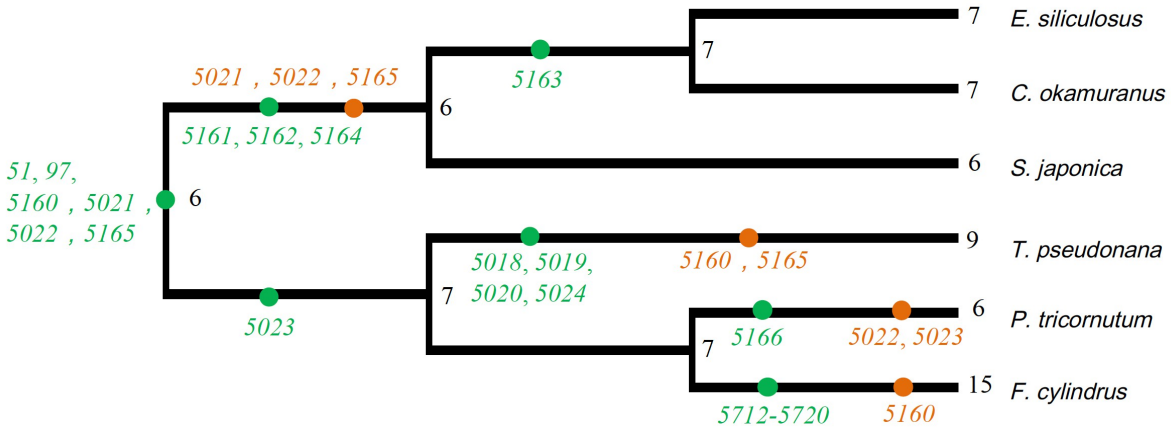

E. siliculosus
C. okamuranus
S. japonica
T. pseudonana
P. tricornutum
F. cylindrus
5021 , 5022 , 5165
5161, 5162, 5164
51, 97,
5160 , 5021 ,
5022 , 5165
5023
5163
5018, 5019,
5020, 5024
5160 , 5165
5166
5022, 5023
5712-5720
5160

fig.3

Figure 4

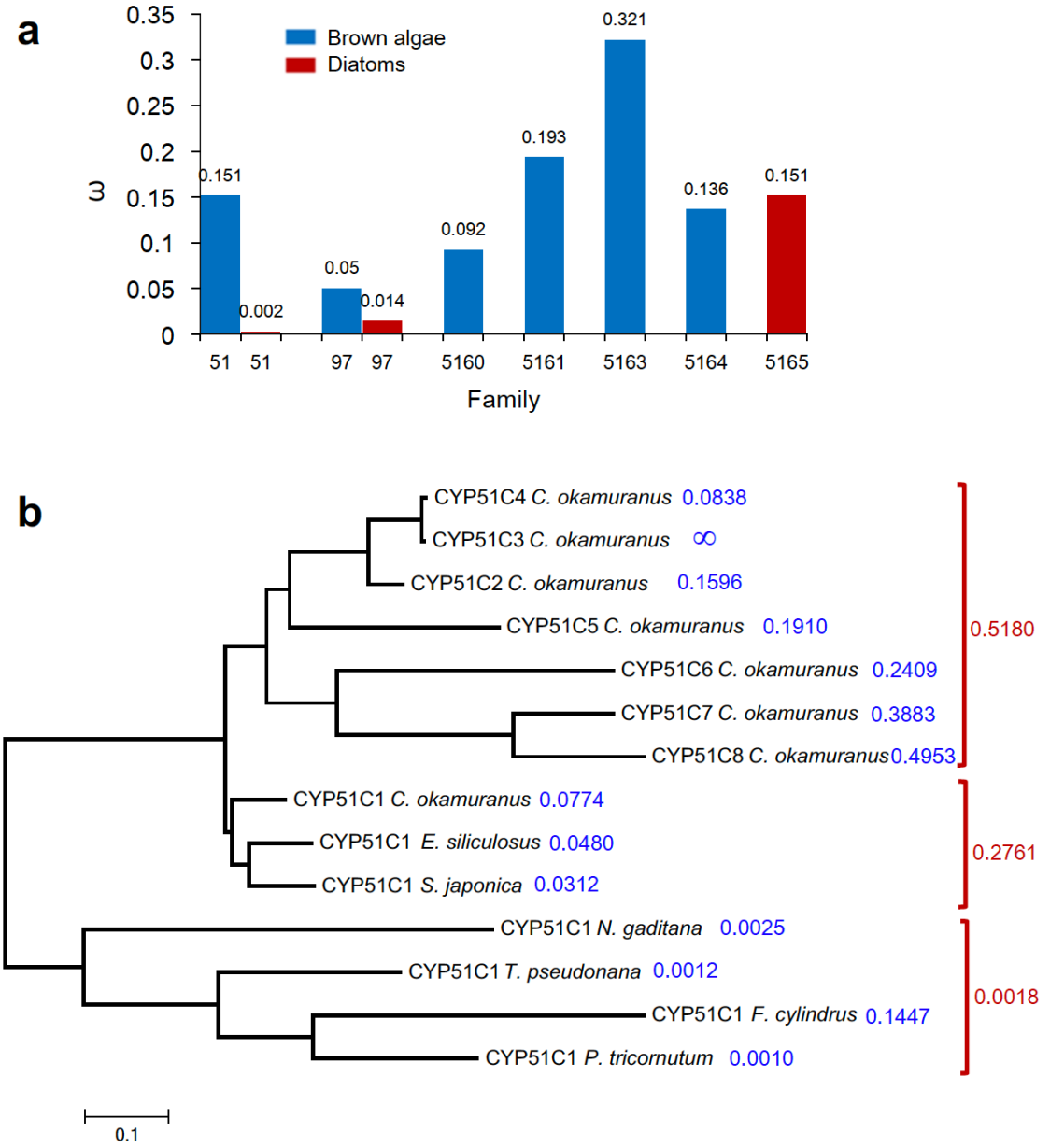

a
Brown algae
Diatoms
ω
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0.151
0.002
0.05
0.014
0.092
0.193
0.321
0.136
0.151
51
51
97
97
5160
5161
5163
5164
5165
Family
b
CYP51C4 C. okamuranus 0.0838
CYP51C3 C. okamuranus ∞
CYP51C2 C. okamuranus 0.1596
CYP51C5 C. okamuranus 0.1910
CYP51C6 C. okamuranus 0.2409
CYP51C7 C. okamuranus 0.3883
CYP51C8 C. okamuranus 0.4953
0.5180
CYP51C1 C. okamuranus 0.0774
CYP51C1 E. siliculosus 0.0480
CYP51C1 S. japonica 0.0312
0.2761
CYP51C1 N. gaditana 0.0025
CYP51C1 T. pseudonana 0.0012
CYP51C1 F. cylindrus 0.1447
CYP51C1 P. tricornutum 0.0010
0.0018
0.1

Figure 5

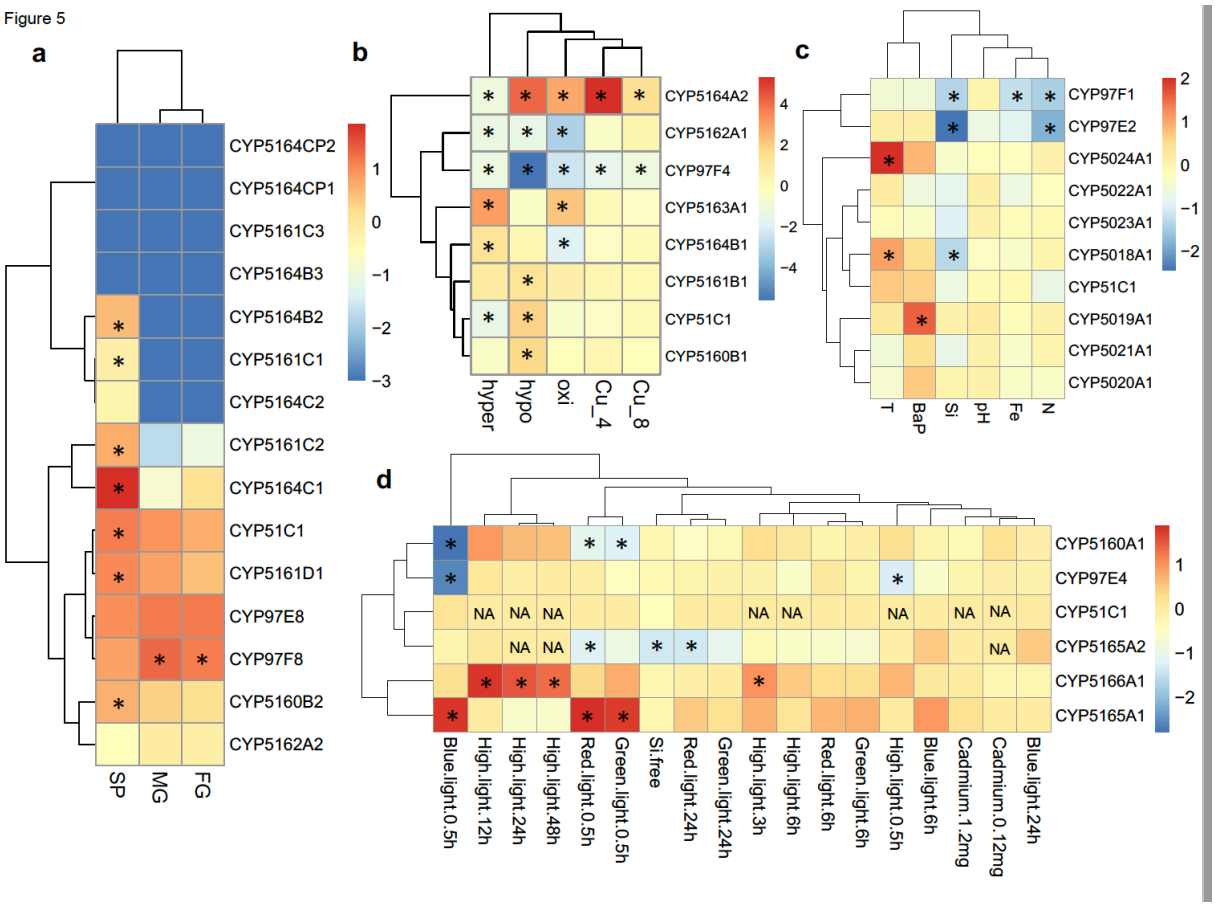
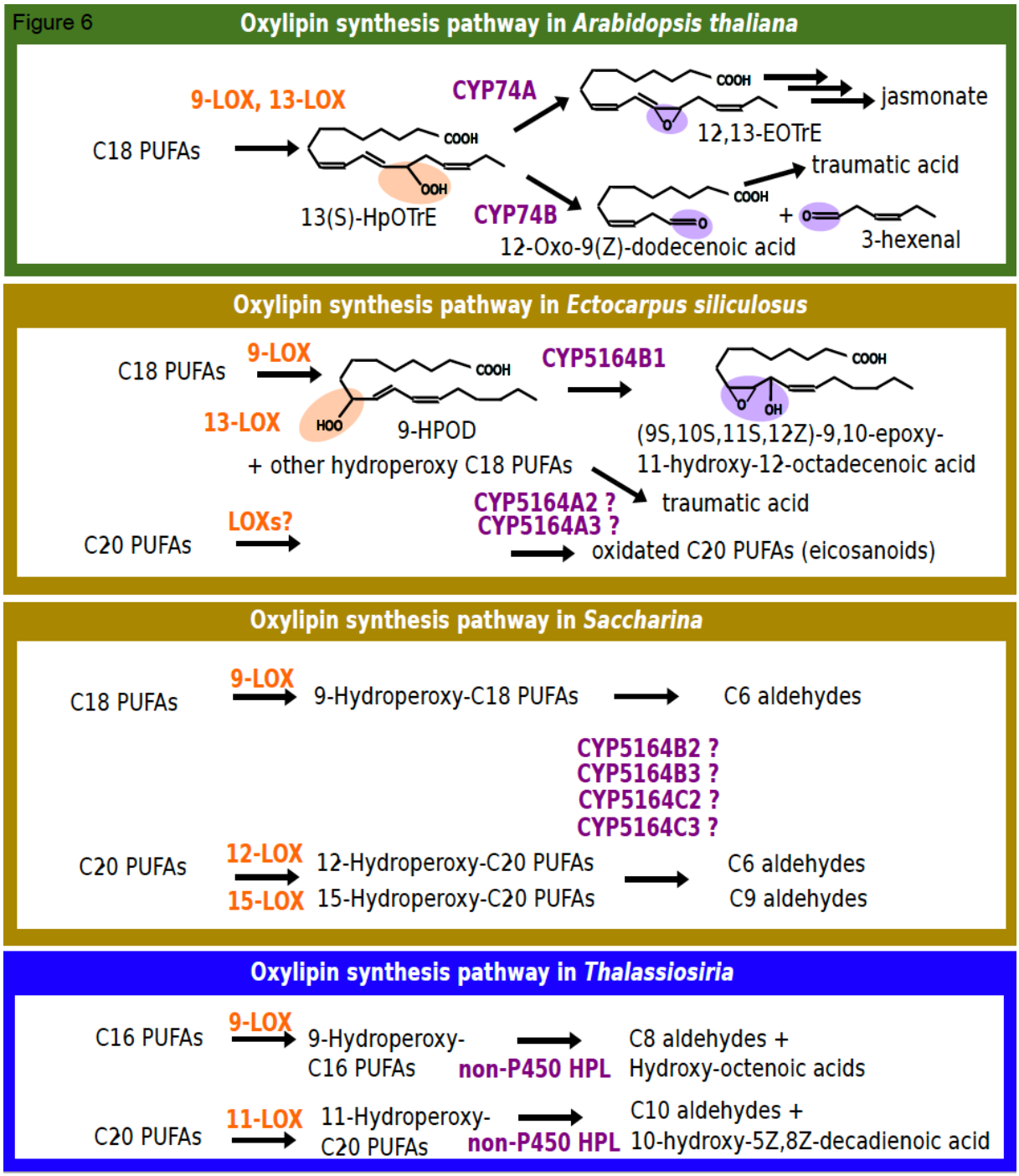

a
CYP5164CP2
CYP5164CP1
CYP5161C3
CYP5164B3
CYP5164B2
CYP5161C1
CYP5164C2
CYP5161C2
CYP5164C1
CYP51C1
CYP5161D1
CYP97E8
CYP97F8
CYP5160B2
CYP5162A2
SP
MG
FG
b
CYP5164A2
CYP5162A1
CYP97F4
CYP5163A1
CYP5164B1
CYP5161B1
CYP51C1
CYP5160B1
hyper
hypo
oxi
Cu_4
Cu_8
c
CYP97F1
CYP97E2
CYP5024A1
CYP5022A1
CYP5023A1
CYP5018A1
CYP51C1
CYP5019A1
CYP5021A1
CYP5020A1
T
BaP
Si
pH
Fe
N
d
CYP5160A1
CYP97E4
CYP51C1
CYP5165A2
CYP5166A1
CYP5165A1
Blue.light.0.5h
High.light.12h
High.light.24h
High.light.48h
Red.light.0.5h
Green.light.0.5h
Si.free
Red.light.24h
Green.light.24h
High.light.3h
High.light.6h
Red.light.6h
Green.light.6h
High.light.0.5h
Blue.light.6h
Cadmium.1.2mg
Cadmium.0.12mg
Blue.light.24h

Figure 6


Oxylipin synthesis pathway in Arabidopsis thaliana
9-LOX, 13-LOX
CYP74A
C18 PUFAs
COOH
OOH
13(S)-HpOTrE
12,13-EOTrE
jasmonate
traumatic acid
CYP74B
12-Oxo-9(Z)-dodecenoic acid
3-hexenal
Oxylipin synthesis pathway in Ectocarpus siliculosus
C18 PUFAs
9-LOX
13-LOX
HOO
9-HPOD
CYP5164B1
OH
(9S,10S,11S,12Z)-9,10-epoxy-
11-hydroxy-12-octadecenoic acid
+ other hydroperoxy C18 PUFAs
CYP5164A2 ?
CYP5164A3 ?
traumatic acid
C20 PUFAs
LOXs?
oxidated C20 PUFAs (eicosanoids)
Oxylipin synthesis pathway in Saccharina
C18 PUFAs
9-LOX
9-Hydroperoxy-C18 PUFAs
C6 aldehydes
CYP5164B2 ?
CYP5164B3 ?
CYP5164C2 ?
CYP5164C3 ?
C20 PUFAs
12-LOX
12-Hydroperoxy-C20 PUFAs
C6 aldehydes
15-LOX
15-Hydroperoxy-C20 PUFAs
C9 aldehydes
Oxylipin synthesis pathway in Thalassiosiria
C16 PUFAs
9-LOX
9-Hydroperoxy-
C16 PUFAs
non-P450 HPL
C8 aldehydes +
Hydroxy-octenoic acids
C20 PUFAs
11-LOX
11-Hydroperoxy-
C20 PUFAs
non-P450 HPL
C10 aldehydes +
10-hydroxy-5Z,8Z-decadienoic acid

Supplementary material

Diversity and evolution of cytochromes P450 in Stramenopiles

Linhong Teng[1,2], Xiao Fan[1], David R. Nelson[3], Wentao Han[1], Xiaowen Zhang[1], Dong Xu[1], Hugues Renault[4], Gabriel V. Markov[5], Naihao Ye[1,2,*]

[1]Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, China, 266071
[2]Function Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, China, 266071
[3]Department of Microbiology, Immunology and Biochemistry, University of Tennessee Health Science Center, 858 Madison Ave. Suite G01, Memphis, TN, USA 38163
[4]Institute of Plant Molecular Biology, CNRS, University of Strasbourg, 67084 Strasbourg, France
[5]Sorbonne Université, CNRS, Integrative Biology of Marine Models (LBI2M), Station Biologique de Roscoff (SBR), 29680 Roscoff, France

*Corresponding author: E-mail, yenh@ysfri.ac.cn; Tel/Fax: +86-532-85830360

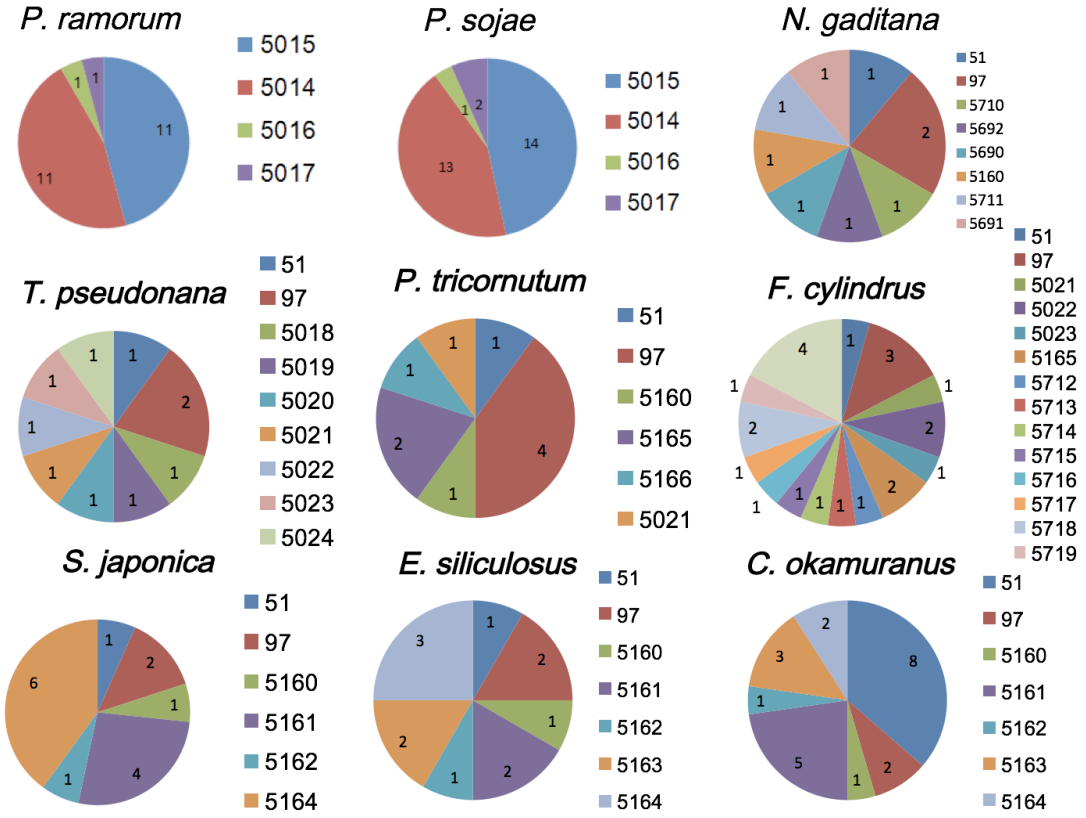


Fig. S1 Statistics of P450 families in each stramenopiles species.

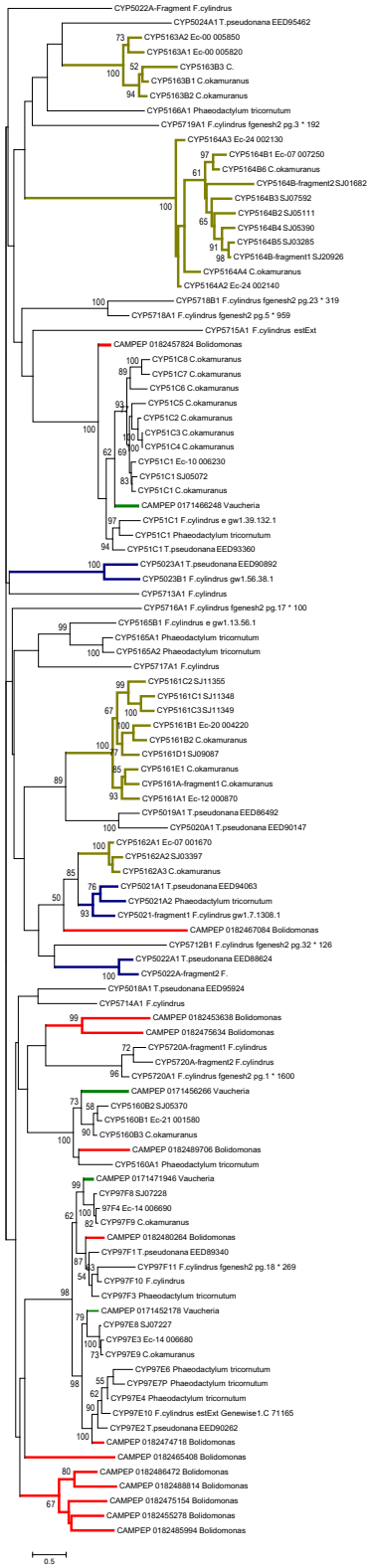


Fig. S2 Phylogenetic tree of CYPs in brown algae, diatoms, *Bolidomonas pacifica* and *Vaucheria litorea*. The ML tree was constructed using MEGA 7.0 with WAG+G model. The dark yellow branches are CYPs only occurring in brown algae. The blue branches are CYPs only occurring in diatoms. The red and green branches correspond to CYPs from *Bolidomonas* and *Vaucheria*, respectively.

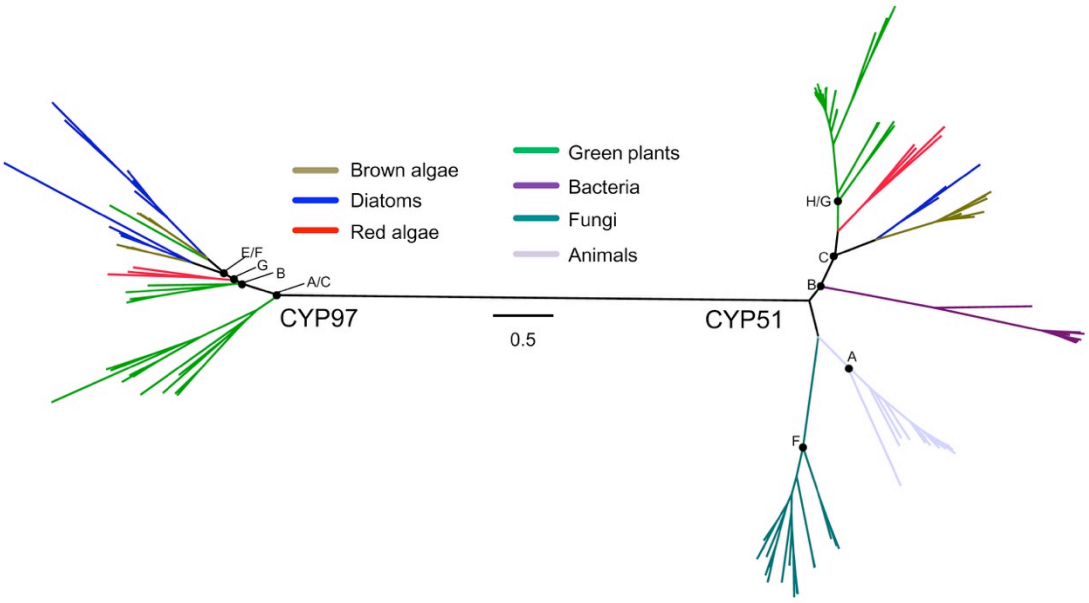


Fig. S3 Phylogeny of brown algae and diatoms CYP51 and CYP97 with their counterparts in animal, plant, bacterial and fungi. The phylogenetic tree was constructed with MEGA 7 and displayed with Figtree, using the alignments of CYPs generated by hmmalign in HMMER program with CYP sequences and PF00067.hmm as input files. 16 animal CYP51s, 10 bacteria CYP51s, 27 plant CYP51s, 21 fungal CYP51s and 16 plant CYP97s were extracted from the CYP homepage http://drnelson.uthsc.edu/CytochromeP450.html.

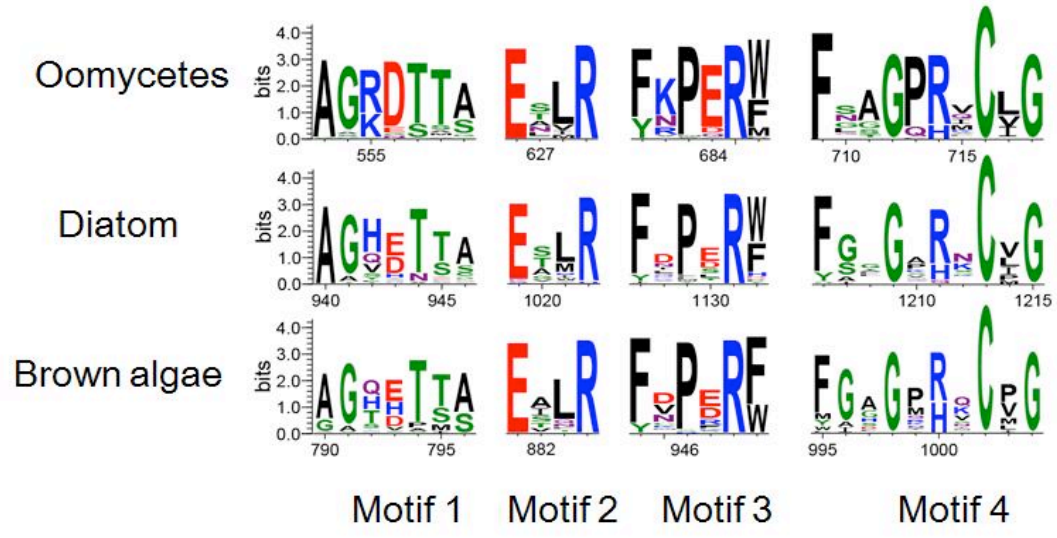


Fig. S4 Sequence logos of the conserved CYP motifs from the three lineages of oomycetes, diatoms and brown algae. Multiple alignments of CYP proteins were performed by aligning them to the PF00067.hmm with HMMER package. The consensus logos were generated by Weblogo http://weblogo.threeplusone.com/create.cgi.

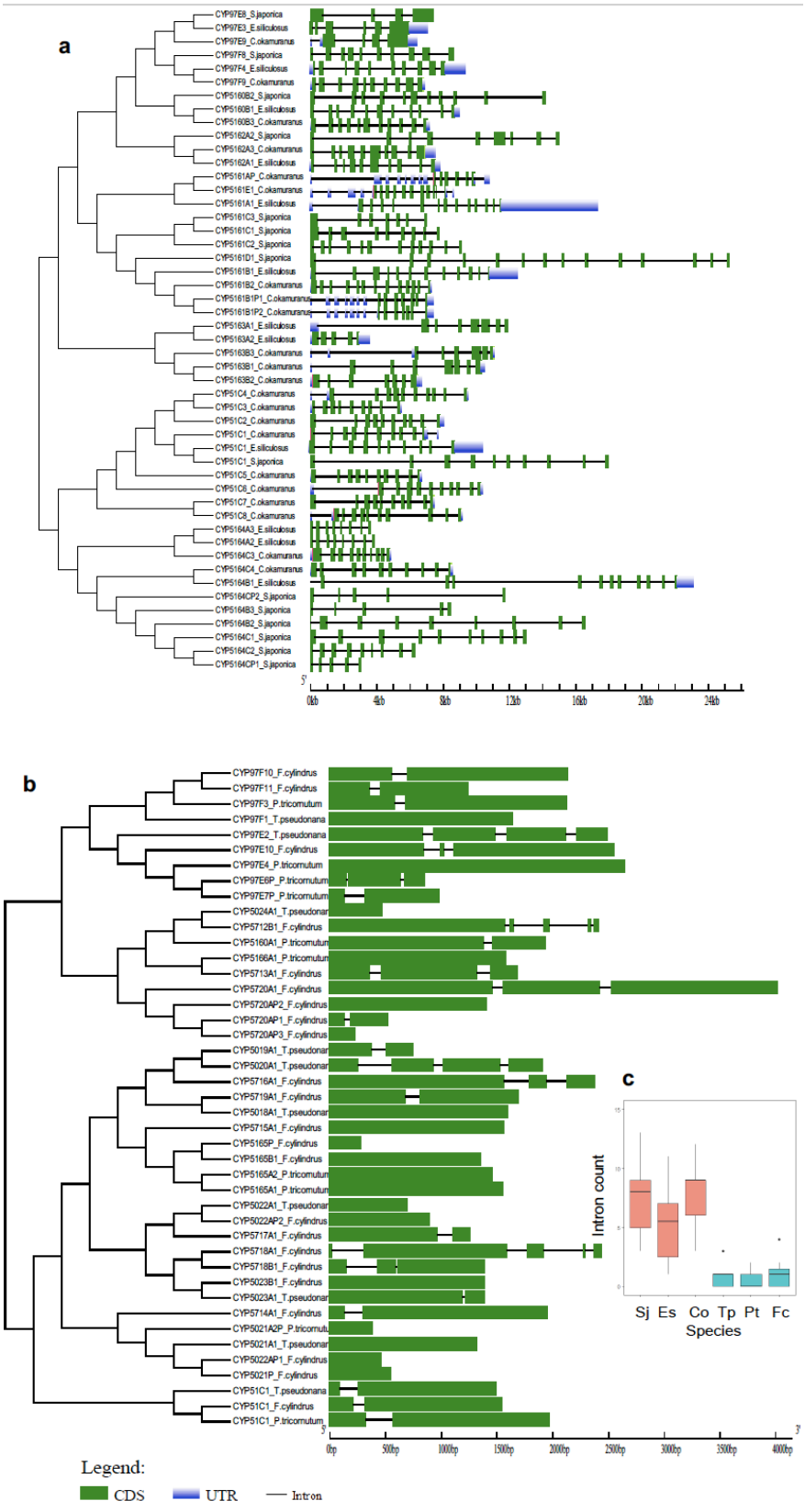
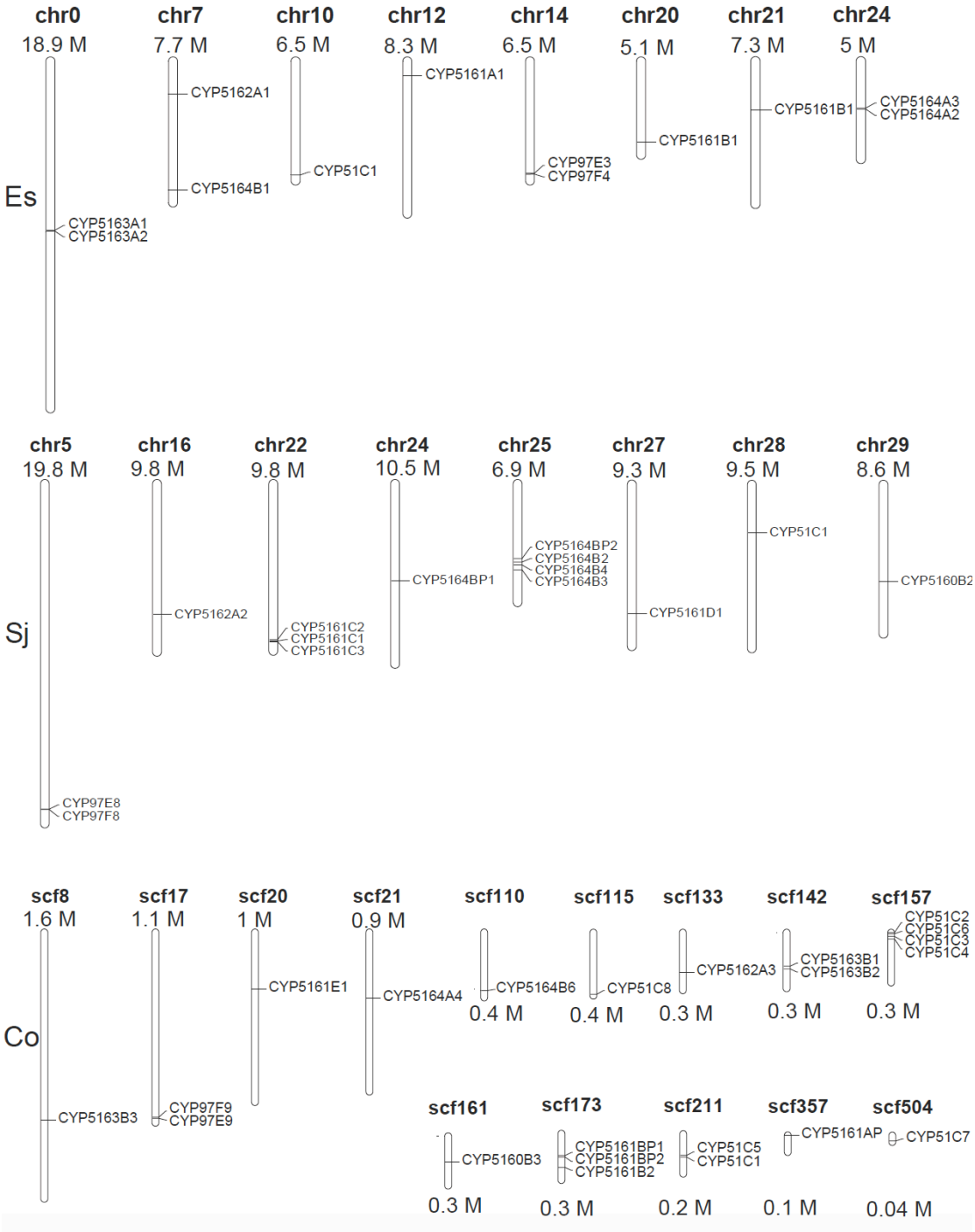


Fig. S5 The gene structure of P450 genes. **a** The gene structure of P450 genes from brown algae. **b** The gene structure of P450 genes from diatoms. **c** The boxplot showing the intron count distribution.


Es
chr0
18.9 M
CYP5163A1
CYP5163A2
chr7
7.7 M
CYP5162A1
CYP5164B1
chr10
6.5 M
CYP51C1
chr12
8.3 M
CYP5161A1
chr14
6.5 M
CYP97E3
CYP97F4
chr20
5.1 M
CYP5161B1
chr21
7.3 M
CYP5161B1
chr24
5 M
CYP5164A3
CYP5164A2
Sj
chr5
19.8 M
CYP97E8
CYP97F8
chr16
9.8 M
CYP5162A2
chr22
9.8 M
CYP5161C2
CYP5161C1
CYP5161C3
chr24
10.5 M
CYP5164BP1
chr25
6.9 M
CYP5164BP2
CYP5164B2
CYP5164B4
CYP5164B3
chr27
9.3 M
CYP5161D1
chr28
9.5 M
CYP51C1
chr29
8.6 M
CYP5160B2
Co
scf8
1.6 M
CYP5163B3
scf17
1.1 M
CYP97F9
CYP97E9
scf20
1 M
CYP5161E1
scf21
0.9 M
CYP5164A4
scf110
CYP5164B6
0.4 M
scf115
CYP51C8
0.4 M
scf133
CYP5162A3
0.3 M
scf142
CYP5163B1
CYP5163B2
0.3 M
scf157
CYP51C2
CYP51C6
CYP51C3
CYP51C4
0.3 M
scf161
CYP5160B3
0.3 M
scf173
CYP5161BP1
CYP5161BP2
CYP5161B2
0.3 M
scf211
CYP51C5
CYP51C1
0.2 M
scf357
CYP5161AP
0.1 M
scf504
CYP51C7
0.04 M

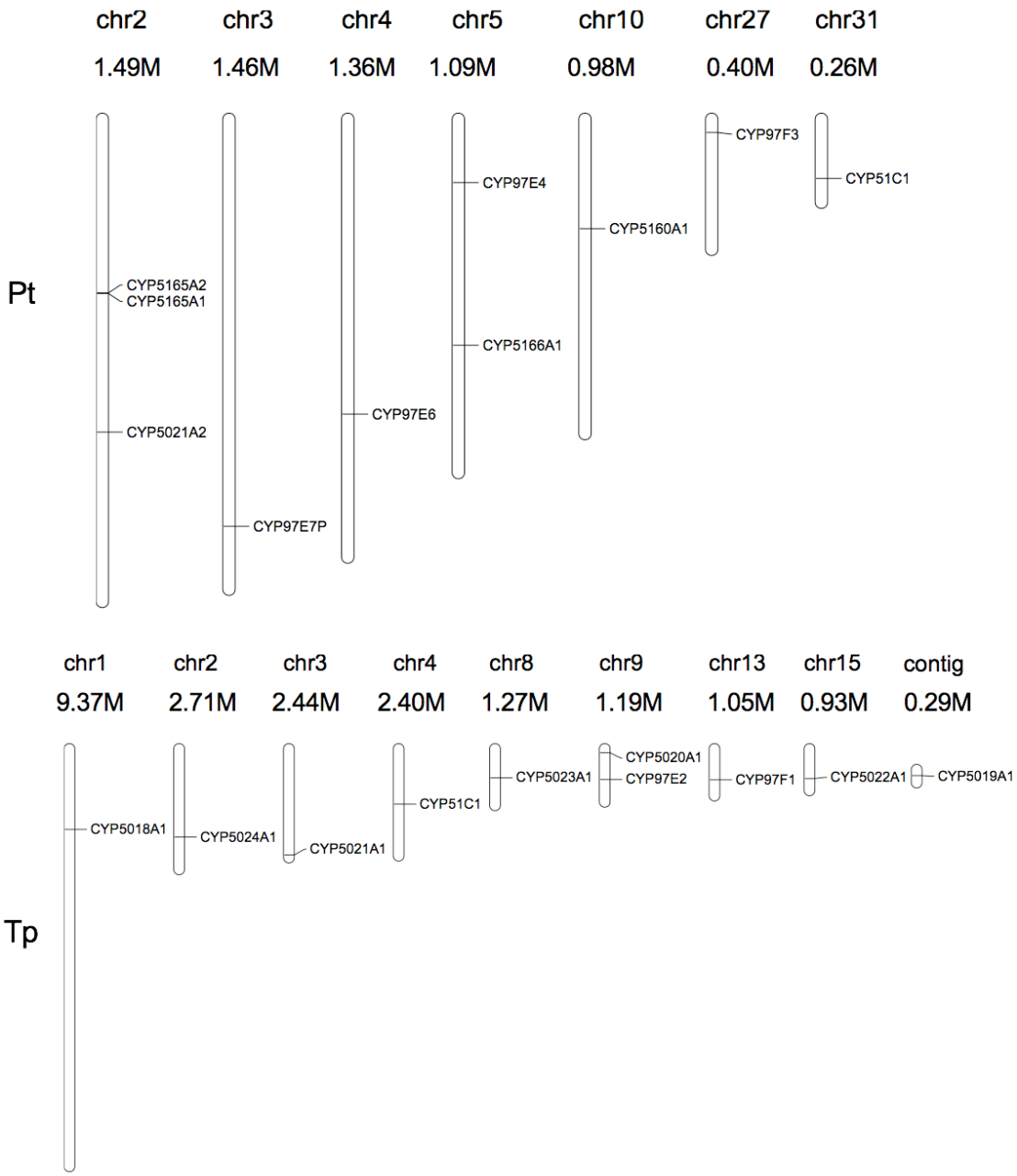


Fig. S6 Chromosome distribution of P450s in brown algae and diatoms. The position of each P450 was mapped to the genome. Es: *E. siliculosus*, Sj: *S. japonica*, Co: *C. okamuranus*, Pt: *Phaeodactylum tricornutum*, Tp: *Thalassiosira pseudonana*

Table S1 Statistics of P450 number in each species.

Species name	No. of P450s	No. of P450 families	No. of P450 subfamilies
Phytophthora ramorum	24	4	17
Phytophthora sojae	30	4	18
Thalassiosira pseudonana	10	9	10
Phaeodactylum tricornutum	10	6	7
Fragilariopsis cylindrus	23	15	17
Saccharina japonica	15	6	8
Ectocarpus siliculosus	12	7	10
Cladosiphon okamuranus	22	7	10
Nannochloropsis gaditana	9	8	9

Table S2 Results of BLAST against 1KP transcriptome database (https://db.cngb.org/onekp) using CYPs only occurring in brown algae (CYP5161, CYP5162, CYP5163 and CYP5164) and CYPs only occurring in diatoms (CYP5021, CYP5022, CYP5023 and CYP5165).

Lineage	Subject Seq-id queried with CYP5161	Identity
brown algae	gnl|onekp|RAPY_scaffold_2003292	47.989
brown algae	gnl|onekp|FIDQ_scaffold_2071904	49.292
brown algae	gnl|onekp|QLMZ_scaffold_2000786	48.611
brown algae	gnl|onekp|JCXF_scaffold_2007092	46.277
brown algae	gnl|onekp|ULXR_scaffold_2004365	48.023
brown algae	gnl|onekp|ULXR_scaffold_2004363	48.023
brown algae	gnl|onekp|VRGZ_scaffold_2008152	46.561
brown algae	gnl|onekp|JCXF_scaffold_2075000	46.348
brown algae	gnl|onekp|JCXF_scaffold_2013398	46.175
brown algae	gnl|onekp|JGGD_scaffold_2003359	47.74
brown algae	gnl|onekp|JGGD_scaffold_2003358	47.74
brown algae	gnl|onekp|QLMZ_scaffold_2011998	47.268
brown algae	gnl|onekp|QLMZ_scaffold_2007179	47.753
brown algae	gnl|onekp|FIKG_scaffold_2006440	45.245
brown algae	gnl|onekp|FOMH_scaffold_2005870	43.333
brown algae	gnl|onekp|FOMH_scaffold_2017792	44.957
brown algae	gnl|onekp|RAPY_scaffold_2084773	62
brown algae	gnl|onekp|RWXW_scaffold_2000875	42.075
brown algae	gnl|onekp|VRGZ_scaffold_2011454	45.278
brown algae	gnl|onekp|ULXR_scaffold_2002547	47.159
brown algae	gnl|onekp|RAPY_scaffold_2084718	85.401
brown algae	gnl|onekp|ULXR_scaffold_2006909	54.077
brown algae	gnl|onekp|ASZK_scaffold_2003291	50
brown algae	gnl|onekp|ASZK_scaffold_2003290	50.235
brown algae	gnl|onekp|ASZK_scaffold_2003289	51.931
brown algae	gnl|onekp|ASZK_scaffold_2003288	51.931
brown algae	gnl|onekp|YRMA_scaffold_2004016	43.066
brown algae	gnl|onekp|YRMA_scaffold_2004013	43.066
brown algae	gnl|onekp|JGGD_scaffold_2009173	53.03
brown algae	gnl|onekp|VYER_scaffold_2004128	51.208
brown algae	gnl|onekp|FIKG_scaffold_2008304	55.376
brown algae	gnl|onekp|ULXR_scaffold_2004364	47.465
brown algae	gnl|onekp|RWXW_scaffold_2011652	53.631
brown algae	gnl|onekp|ASZK_scaffold_2003292	55.122
brown algae	gnl|onekp|QDTV_scaffold_2005010	57.436
brown algae	gnl|onekp|OGZM_scaffold_2006939	57.436
brown algae	gnl|onekp|QLMZ_scaffold_2009369	51.22
brown algae	gnl|onekp|VRGZ_scaffold_2009288	48.718
brown algae	gnl|onekp|ULXR_scaffold_2004367	45.815
brown algae	gnl|onekp|ASZK_scaffold_2003335	49.215
brown algae	gnl|onekp|ASZK_scaffold_2003334	49.215
brown algae	gnl|onekp|YRMA_scaffold_2019025	49.457
brown algae	gnl|onekp|LIRF_scaffold_2099847	42.466
brown algae	gnl|onekp|RWXW_scaffold_2004548	44.693
brown algae	gnl|onekp|VRGZ_scaffold_2007856	45.078
brown algae	gnl|onekp|RAPY_scaffold_2087423	54.658
brown algae	gnl|onekp|VYER_scaffold_2004127	49.669
brown algae	gnl|onekp|QLMZ_scaffold_2013765	43.655
brown algae	gnl|onekp|LDRY_scaffold_2010562	42.553
brown algae	gnl|onekp|RWXW_scaffold_2000877	45.455
brown algae	gnl|onekp|YRMA_scaffold_2003902	47.407
brown algae	gnl|onekp|LDRY_scaffold_2018822	47.407
brown algae	gnl|onekp|RWXW_scaffold_2004547	45.926
brown algae	gnl|onekp|RWXW_scaffold_2000876	42.958
brown algae	gnl|onekp|LDRY_scaffold_2012678	43.662
brown algae	gnl|onekp|VRGZ_scaffold_2010578	49.057
brown algae	gnl|onekp|LDRY_scaffold_2019529	49.541
brown algae	gnl|onekp|FIKG_scaffold_2003296	46.903
brown algae	gnl|onekp|QLMZ_scaffold_2008333	46.903
brown algae	gnl|onekp|HFIK_scaffold_2061728	51.485
brown algae	gnl|onekp|JGGD_scaffold_2013374	56.25
brown algae	gnl|onekp|FIKG_scaffold_2003295	47.115

Lineage	Subject Seq-id queried with CYP5162	Identity
brown algae	gnl|onekp|FSQE_scaffold_2048840	86.905
brown algae	gnl|onekp|JCXF_scaffold_2002771	75.13
brown algae	gnl|onekp|LIRF_scaffold_2013308	75.449
brown algae	gnl|onekp|VYER_scaffold_2015145	71.667
brown algae	gnl|onekp|RWXW_scaffold_2007060	74.556
brown algae	gnl|onekp|YRMA_scaffold_2103644	71.667
brown algae	gnl|onekp|FIKG_scaffold_2015075	75.301
brown algae	gnl|onekp|QLMZ_scaffold_2013683	74.126
brown algae	gnl|onekp|FOMH_scaffold_2015399	74.809
brown algae	gnl|onekp|HFIK_scaffold_2063485	72.18
brown algae	gnl|onekp|JGGD_scaffold_2076440	80
brown algae	gnl|onekp|VRGZ_scaffold_2081428	83.178
brown algae	gnl|onekp|ULXR_scaffold_2061008	80.556

Lineage	Subject Seq-id queried with CYP5163	Identity
brown algae	gnl|onekp|VRGZ_scaffold_2020794	56.355
brown algae	gnl|onekp|QLMZ_scaffold_2006153	50.265
brown algae	gnl|onekp|RWXW_scaffold_2008007	43.478
brown algae	gnl|onekp|ASZK_scaffold_2002108	52.464
brown algae	gnl|onekp|VYER_scaffold_2002462	41.905
brown algae	gnl|onekp|YRMA_scaffold_2018088	43.175
brown algae	gnl|onekp|VYER_scaffold_2011777	41.495
brown algae	gnl|onekp|OGZM_scaffold_2006379	43.048
brown algae	gnl|onekp|OGZM_scaffold_2006378	43.048
brown algae	gnl|onekp|LDRY_scaffold_2000724	42.276
brown algae	gnl|onekp|JCXF_scaffold_2013906	52.113
brown algae	gnl|onekp|RWXW_scaffold_2073915	41.139
brown algae	gnl|onekp|ULXR_scaffold_2069574	60.847
brown algae	gnl|onekp|ASZK_scaffold_2002107	58.201
brown algae	gnl|onekp|FOMH_scaffold_2012469	45.872
brown algae	gnl|onekp|VYER_scaffold_2010815	48.37
brown algae	gnl|onekp|HFIK_scaffold_2005799	43.111
brown algae	gnl|onekp|FIKG_scaffold_2004819	49.451
brown algae	gnl|onekp|LDRY_scaffold_2004052	50.543
brown algae	gnl|onekp|LDRY_scaffold_2000723	46.842
brown algae	gnl|onekp|RWXW_scaffold_2005981	47.895
brown algae	gnl|onekp|RWXW_scaffold_2005978	47.368
brown algae	gnl|onekp|RWXW_scaffold_2005979	47.368
brown algae	gnl|onekp|APTP_scaffold_2010405	46.701
brown algae	gnl|onekp|LDRY_scaffold_2000727	48
brown algae	gnl|onekp|LIRF_scaffold_2098123	42.439
brown algae	gnl|onekp|JGGD_scaffold_2072503	55.455
brown algae	gnl|onekp|HFIK_scaffold_2001187	54.867
brown algae	gnl|onekp|FSQE_scaffold_2040927	57.282
brown algae	gnl|onekp|LDRY_scaffold_2011830	55.34
brown algae	gnl|onekp|QDTV_scaffold_2009548	48.819
brown algae	gnl|onekp|QDTV_scaffold_2009547	50.413
brown algae	gnl|onekp|HFIK_scaffold_2057736	50
brown algae	gnl|onekp|FSQE_scaffold_2044923	61.765

Lineage	Subject Seq-id queried with CYP5164	Identity
brown algae	gnl|onekp|JCXF_scaffold_2010654	80.42
brown algae	gnl|onekp|ULXR_scaffold_2006209	79.856
brown algae	gnl|onekp|ULXR_scaffold_2005872	75.887
brown algae	gnl|onekp|JCXF_scaffold_2010655	77.444

Lineage	Subject Seq-id queried with CYP5021	Identity
brown algae	gnl|onekp|FSQE_scaffold_2048840	47.312
brown algae	gnl|onekp|LIRF_scaffold_2013308	46.597
brown algae	gnl|onekp|JCXF_scaffold_2002771	41.204
brown algae	gnl|onekp|VYER_scaffold_2015145	43.478
brown algae	gnl|onekp|RWXW_scaffold_2007060	43.243
brown algae	gnl|onekp|YRMA_scaffold_2103644	43.478
brown algae	gnl|onekp|FIKG_scaffold_2015075	44.505
brown algae	gnl|onekp|QLMZ_scaffold_2013683	44.805
brown algae	gnl|onekp|FOMH_scaffold_2015399	45
brown algae	gnl|onekp|HFIK_scaffold_2063485	45
brown algae	gnl|onekp|JGGD_scaffold_2076440	42.623
brown algae	gnl|onekp|VRGZ_scaffold_2081428	42.623

Lineage	Subject Seq-id queried with CYP5022	Identity
Cryptophyta	gnl|onekp|IRZA_scaffold_2061631	42.424

Lineage	Subject Seq-id queried with CYP5165	Identity
Chrysophyceae	gnl|onekp|EBWI_scaffold_2034733	44.586
Haptophyceae	gnl|onekp|NMAK_scaffold_2038424	43.75
Haptophyceae	gnl|onekp|LLEN_scaffold_2044944	43.75

Lineage	Subject Seq-id queried with CYP5023	Identity
Cryptophyta	gnl|onekp|ROZZ_scaffold_2062059	28.904

Table S3 CYP list, showing the CYP name, protein length, subcellular location, and transmembrane helix number (TH)

CYP name	Protein length	Subcellular location	TH
Thalassiosira pseudonana			
51C1	446	ER	0
97E2	736	Plastid	0
97F1	546	ER	0
5018A1	537	ER	0
5019A1	625	Cytoplasm, ER, Nucleus	0
5020A1	463	ER	2
5021A1	418	ER ,Mitochondrion	0
5022A1	488	ER	0
5023A1	457	ER	0
5024A1	530	ER	0
Fragilariopsis cylindrus			
5717A1	374	ER	0
97F10	627	Plastid	0
5165B1	453	ER	0
5022AP2	299	ER	0

5022AP1	156	ER, Mitochondrion	0
5720A1	1273	ER	0
5718B1	392	ER	0
5021P1	183	ER	0
97E10	741	Plastid	2
5165P1	94	ER, Mitochondrion	0
97F11	382	Cytoplasm, ER	1
5718A1	517	ER	0
5719A1	522	ER	1
5712B1	579	Plastid	0
51C1	479	Plastid	1
5716A1	655	Cytoplasm,ER, nucleus	0
5720AP1	158	Cytoplasm, ER	0
5720AP3	76	ER	0
5023B1	463	ER	0
5720AP2	144	ER	0
5714A1	597	ER	0
5715A1	520	ER	2
5713A1	488	ER	0
Phaeodactylum tricornutum			
51C1	482	ER	1
97F3	538	ER	0
5160A1	565	ER	1
5166A1	528	ER	0
97E4	769	ER	0
97E6	345	ER	0
97E7P	266	ER, Mitochondrion	0
5021A2	634	ER	0
5165A1	488	ER	2
5165A2	487	ER	1
Saccharina japonica			
51C1	485	ER	1
97E8	756	ER, Mitochondrion	0
97F8	633	Plastid	0
5160B2	571	ER	0
5161C1	471	ER	2
5161C2	485	ER	2
5161C3	348	ER	1
5161D1	574	ER	0
5162A2	615	ER, Mitochondrion	1
5164B2	433	ER	0
5164B3	359	Cytoplasm, ER	0
5164B4	480	Cytoplasm, ER	0
5164B5	382	Cytoplasm, ER	0
5164BP1	212	ER	0
5164BP2	173	Mitochondrion	0
Ectocarpus siliculosus			
5162A1	659	Cytoplasm, ER, Mitochondrion	2
5164B1	536	Cytoplasm, ER	0
51C1	520	ER	3
5161A1	530	ER	0
97E3	774	Cytoplasm, ER	0
97F4	627	Plastid	0
5161B1	562	ER	1
5160B1	560	ER	1
5164A3	300	ER	0
5164A2	286	ER	0
5163A2	552	Plastid	0
5163A1	564	Plastid	1
Cladosiphon okamuranus			

5163B3	398	Cytoplasm, ER	0
97F9	638	Plastid	1
97E9	763	ER	0
5161E1	345	ER	0
5164A4	546	Cell membrane, ER	0
5164B6	532	Plastid	0
51C8	494	ER	1
5162A3	728	ER, Mitochondrion	2
5163B1	527	ER	0
5163B2	463	Plastid	1
51C2	490	ER	1
51C6	491	ER	1
51C3	414	ER	0
51C4	490	ER	1
5160B3	558	ER	0
5161BP1	229	ER	0
5161BP2	229	ER	0
5161B2	514	ER	0
51C5	492	ER	1
51C1	483	ER	2
5161AP1	214	ER	0
51C7	492	ER	1
Nannochloropsis gaditana			
97F5	540	ER	0
CYP5710A1	1221	ER	0
CYP5692A2	527	ER	1
51C1	507	ER	0
CYP5690A1	518	ER	1
CYP5160C2	579	ER	0
CYP5711A1	375	ER	0
CYP97E5	699	ER	0
CYP5691A2P	163	ER	0

Table S4 Sequence identity between brown algal CYP51.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1.CYP51C1_Sj										
2.CYP51C1_Es	0.817									
3.CYP51C8_Co	0.621	0.588								
4.CYP51C2_Co	0.781	0.723	0.611							
5.CYP51C6_Co	0.649	0.618	0.582	0.641						
6.CYP51C3_Co	0.651	0.616	0.51	0.761	0.547					
7.CYP51C4_Co	0.757	0.698	0.603	0.889	0.645	0.838				
8.CYP51C5_Co	0.719	0.666	0.591	0.707	0.627	0.594	0.699			
9.CYP51C1_Co	0.86	0.813	0.619	0.787	0.653	0.671	0.761	0.731		
10.CYP51C7_Co	0.646	0.608	0.765	0.646	0.6	0.534	0.623	0.63	0.652	

Table S5 Statistics of Codeml calculation.

Family	M0 (one-ratio)		Site model -lnL		2ΔlnL	p-value
	ω	-lnL	M7	M8		
Brown algae CYP51	0.1509	6911.09	6849.41	6848.71	1.4	
Diatom CYP51	0.00211	4126.37	4083.87	4083.59	0.56	
Brown algae CYP97	0.05018	8327.44	8067.5	8046.25	42.5	<0.001
Diatom CYP97	0.01406	8203.9	8055.24	8054.17	2.14	
Brown algae CYP5160	0.09232	4482.55	4437.14	4433.33	7.62	<0.001
Brown algae CYP5161	0.19334	4391.33	4326.2	4326.19	0.02	
Brown algae CYP5163	0.32151	3669.85	3651.38	3651.11	0.54	
Brown algae CYP5164	0.13675	5847.88	5772.83	5772.82	0.02	
Diatom CYP5165	0.15071	4402.62	4375.29	4373.11	4.36	
CYP51 branch site model	-lnL1	-lnL2	2ΔlnL			
Foreground C3-C7	10161.6	10161.5	0.2			
Foreground C3	10163.5	10163.4	0.2			