Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Pole-to-pole LOV-domain receptor diversity points to an early stramenopile origin and global dominance of Aureochromes

Melissa Misir

melissa.misir@uni-konstanz.de

Universität Konstanz https://orcid.org/0009-0003-1051-1082

Peter Kroth

University of Konstanz https://orcid.org/0000-0003-4734-8955

Sylke Wohlrab

Article

Keywords:

Posted Date: March 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9010570/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Abstract

Light regimes are a fundamental environmental cue for marine organisms, yet molecular adaptations of photoreceptors to the extreme seasonality of polar regions remain poorly understood. We demonstrate that the diversity of candidate blue-light-sensing LOV-domain proteins increases towards higher latitudes. We identify Aureochromes, blue-light-sensitive transcription factors unique to stramenopiles, as the most frequent LOV-domain receptors across ocean latitudes. Phylogenetic reconstructions reveal that Aureochromes diversified early in the stramenopile lineage, originating in a heterotrophic ancestor before the acquisition of photosynthesis. This challenges the assumption that Aureochromes are restricted to phototrophs, suggesting an ancestral role in spectral light signaling. Despite strong latitudinal shifts in light quality and photoperiod, diatom Aureochrome domain architecture remains conserved, implying adaptation through regulatory tuning (e.g. altered expression dynamics or post-translational control). Together, our results emphasise an importance of spectral-light sensing in polar oceans and link the widespread dominance of Aureochromes to their ancient origin and conservation.

Introduction

The fact that blue wavelengths penetrate seawater more deeply than any other part of the visible spectrum characterizes our blue planet. For marine organisms, blue light accordingly provides reliable information about depth, turbidity, diurnal and annual rhythmicity, and has therefore become a central cue that regulates essential physiological and behavioral processes [1]. Moreover, the underwater light field fuels photosynthetic organisms which constantly have to adjust to changes in light quality and quantity due to e.g. mixing or light-absorbing particles [2,3]. In addition to the ability to quantify light by absorbing it through the photosynthetic machinery, phytoplankton species have specific photoreceptors that recognize a defined range of wavelengths and light intensities in the water column [3].

The dominance of blue light in the water column may be the reason why algae, including phytoplankton, possess a larger number of algal-specific blue light photoreceptors, not found in land plants [4]. Likewise, the phylogenetically diverse group of photosynthetic Stramenopiles harbors a distinctive class of blue-light receptors referred to as Aureochromes (AUREOs) [5]. AUREOs contain a DNA-binding bZIP-domain in the N-terminal region of the protein, followed by the C-terminal, light-sensing flavin-binding LOV-domain (Light-Oxygen-Voltage domain). This architecture makes AUREOs being blue light sensitive transcription factors. In contrast to other photoreceptors that after activation initiate a signaling cascade, AUREOs after activation directly interact with promoters, allowing for rapid responses [5–7]. Gene regulation occurs, after dimerizing the bZIP-domains of two AUREO proteins and binding to a promoter of target genes, by a blue light stimulus [8].

AUREOs have been particularly well-studied in diatoms, which are photoautotrophic algae belonging to the Stramenopiles. These algae are contributing significantly to global primary production [9,10]. So far, four paralogues of AUREO proteins have been described in the model diatom *Phaeodactylum*

tricornutum: AUREO 1a, 1b, 1c, and 2 [11]. The expression of some AUREOs is regulated with a diurnal pattern (PtAureo 1a and 1c), whereas others show light induced expression (PtAureo1b), while PtAureo2 is only weakly regulated [8]. Transcriptomic studies of *P. tricornutum* revealed that PtAureo1a is involved in the regulation of the cell cycle, light acclimation, the diatom clock [12], and an overall regulation of gene transcription [7,13].

Besides AUREOs, diatoms possess further well-described photoreceptors. Diatom phytochromes (DPH) consist of a C-terminal kinase-like domain and a PAS (Per-ARNT-Sim) domain, a GAF (cGMP-specific phosphodiesterase, adenylyl cyclases and FhlA) domain, and PHY (phytochrome-specific) domain, forming the N-terminal photosensory module [14]. DPHs are responding to red and far-red light (R/FR) and regulate gene expression via phosphorylation [15]. Furthermore, diatoms possess cryptochromes, blue-light photoreceptors that are involved in photoprotection and circadian regulation [16]. Cryptochromes consist of a Photolyase Homology Region (PHR) with a variable C-terminal region (CCE) [17]. Phototropins are found in green algae and land plants only; they possess 2 LOV-domains named LOV1 and LOV2 and a kinase domain for initiating a signal cascade. After blue light activation, phototropins trigger downstream processes like phototropism [18]. Finally, heliorhodopsins have recently been discovered in numerous species including diatoms [19].

Several further proteins of unknown function have been identified in algae that possess a LOV-domain combined with other domains [20]. Such proteins include combinations like LOV-Ef-hand, LOV-HSF (heat shock factor), and helmchromes that possess RGS-LOV-domains (Regulator of G protein Signaling). The high variation of LOV-domain containing proteins shows that marine unicellular planktonic organisms harbor a diversity of possible light-sensitive regulatory mechanisms to be able to thrive in changing light environments [20].

While AUREOs from a few model organisms that live in non-polar regions have been studied in detail [5,7,21], much less is known about AUREOs from polar regions [3]. There, apart from low temperatures, light quality and quantity conditions can be very extreme. Light conditions vary from day/night cycles to seasonal cycles, with polar nights as an extreme form of light shortage and continuous summer daylight as the opposite [22]. The fact that diatoms are nevertheless prevalent in these regions [23], indicates that they have successfully adapted to respective conditions. Diatoms can even thrive in particular habitats like brine channels of sea ice [24].

In this study, we investigate how global light regimes influence the diversity and evolution of LOV-domain photoreceptors in marine unicellular eukaryotes. LOV-domains typically contain a canonical photosensory motif (summarised as GXNCRFLQG), with a conserved cysteine linked to the traditional LOV photocycle [25]. This motif provides strong sequence-based evidence consistent with blue-light-responsive regulation, but photochemical activity cannot be inferred from homology alone [25]. Indeed, further LOV-motif variants lacking the conserved cysteine have been shown to initiate blue-light-triggered

responses via non-canonical flavin (e.g. photoreduction) [26,27]. Consequently, we refer to LOV-domain proteins detected as blue-light photoreceptor candidates that require targeted photochemical validation.

In summary, our comparative analyses of LOV-domain-containing proteins indicate an increase in candidate blue-light photoreceptor diversity and repertoire towards lower latitudes. Across all investigated samples, AUREOs were found to be the most frequently detected LOV-domain containing blue-light photoreceptors. By resolving the AUREO phylogeny, we demonstrate that AUREOs diversified early within the stramenopiles, with a strong conservation of the blue-light-receiving domain and a more pronounced diversification of the transcription factor domain.

Material and Methods

Culture conditions and RNA sequencing

The Southern Ocean diatom isolates *Odontella* sp., *Thalassiosira* sp. (2019, Station 35, PS117, 69° 04' S, 17° 19' W), *Chaetoceros sp.* (2004, Polarstern EIFEX ANT-XXI/3, 49° 36S, 0° 05E, open waters of Atlantic Southern Ocean), *Pseudo-Nitzschia* sp. (2022, 51° 28' 41.7" S, 49° 11' 17.8" W), and *Fragilariopsis sp.* were used for this study. All strains were collected outside national jurisdiction (high seas or Antarctic Treaty area) and are therefore not subject to Nagoya Protocol ABS requirements. The strains were cultured in Tissue Culture Flask in Provasoli's enriched F/2 medium [28] with 250 mM $NaSiO_3x5H_2O$ and 5 mM H_2SeO_3 in 34‰ sea salt content. Cells were exposed to a 16:8 light:dark photoperiod at 50 µmol photons m^{-2} s^{-1} and 4°C. Since some AUREOs follow a diel pattern, RNA samples were taken one hour before, 2 hours after, and 6 hours after the onset of the light cycle for Long-read RNA Sequencing. For sampling, aliquots of 15 mL were collected from individual cultures, and subsequently pooled. Cells were harvested by filtration through polycarbonate membrane filters (Merck KGaA, Darmstadt, Germany) with varying pore sizes based on the cell size of the diatom. *Odontella* sp., *Chaetoceros* sp., and *Thalassiosira* sp. were filtered through 5 µm pore size filters while *Pseudo-Nitzschia* sp. and *Fragilariopsis sp.* were filtered through 2 µm and 1 µm pore sizes respectively. The cells were rinsed off the filters with 1 mL TriReagent (Sigma-Aldrich, Steinheim, Germany) and RNA isolation was performed as described in detail by [29].

RNA samples of *Odontella* sp., *Thalassiosira* sp., *Fragilariopsis sp.*, and *Pseudo-Nitzschia* sp. were sequenced by PacBio® sequel II long-read isoform sequencing (ISO-seq) at Novogene (Munich, Germany). The RNA sample of *Chaetoceros debilis* was sequenced using Illumina Nextseq 2000 (Illumina, San Diego, CA, USA) using P3 Reagents kit (2x 150 cycles). The library for Illumina Sequencing was generated using Illumina Stranded mRNA Prep Ligation Kit (Illumina, San Diego, CA, USA) according to the manufacturer’s protocol.

Data collection and raw data processing

The “Sea of Change” metatranscriptome data were obtained from the Joint Genome Institute (JGI) (JGI Proposal ID 532/300780, doi: 10.46936/10.25585/60000951). Protein sequences from genome and transcriptome data were retrieved from the EukProt database (https://evocellbio.com/eukprot/) [30,31]. Additional, already assembled transcriptome sequences of *Thalassiosira* spp., *Thalassiosira minima*, *Minidiscus* spp., *Minidiscus comicus*, *Minidiscus spinulatus*, and *Minidiscus variabilis* were collected from Guajardo et al. [32] and of *Chaetoceros debilis* as well as *Fragilariopsis kerguelensis* from Beszteri et al. [33]. Sequences of marine stramenopile species were taken from Labarre et al. [34], Illumina raw sequences were cleaned with the BBmap (v39.01) [35] and assembled with Trinity [36]. For ISO-seq long reads, raw reads were cleaned with Lima (v2.9.0) [37], and refined and clustered with Iso-Seq (v4.0.0 [38]). Samtools (v.1.16.1) [39] was used to convert .bam files into final fasta files. TransDecoder (v5.7.0) was used to determine open reading frames and translations for all nucleotide sequences [40].

Identification of LOV-domain containing proteins

The presence of LOV-domains in the retrieved sequences was screened using a custom-built hidden Markov model (profile HMMs, http://hmmer.org), built from the published LOV-domain sequence alignment of Coesel et al. [20]. All transcripts that were found to encode a LOV-domain, were further screened for additional domains, including bZIP-domains, by using the PFAM Profile HMM collection and motif-associated significance cut-offs [41]. Sequences containing an N-terminal bZIP-domain followed by a C-terminal LOV-domain were defined as AUREOs.

Analysis of metatranscriptome data

The taxonomy of transcripts with additional domains besides the LOV-domain was determined with MMseq2 [42] and a custom reference database containing revised sequences of the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) [43]. Sequences assigned as microeukaryotes were kept for the analysis of the diversity of LOV-domain encoding transcripts. To test whether LOV-domain diversity is latitudinally structured independently of background taxonomic diversity and sequencing depth, we fitted a Generalised Additive Model (GAM) with LOV-domain richness as the response variable with the R package mgcv and used residual diagnostics to validate model assumption [44]. Based on our hypothesis that biogeographic gradients may influence blue-light sensory diversity, we modelled latitude using a smooth spline to capture potential nonlinear patterns. As LOV-domain richness may be related to the overall taxonomic complexity of the community, we included the Shannon Index (order-level taxonomic diversity) as a linear covariate. The respective order-level clade counts were created from the MMseq2 taxonomy report files with Pavian [45]. Finally, we added the log-transformed total number of contigs per sample as an offset to adjust for differences in sequencing depth across samples. This enabled us to determine the influence of latitude on LOV-domain diversity while accounting for community diversity and sequencing effort.

Phylogenetic reconstruction and duplication history of AUREOs

We reconstructed the evolutionary history of AUREOs across the stramenopiles by identifying their taxonomic distributions, inferring their phylogenetic relationships and approximating gene duplication events using reconciliation models. Specifically, we screened all stramenopile proteomes from the EukProt database v3 [30] for AUREO candidates as described above. We inferred the species phylogeny for all species with AUREOs using OrthoFinder v3.1.2 [46]. The topology of the obtained species tree was compared to the published reference phylogeny by Cho et al. [47] and the tree was rooted accordingly. The gathered AUREO sequences were aligned using MAFFT v7.490 and a high-accuracy iterative refinement approach (L-INS-i) [48]. Poorly aligned, gap-rich regions were trimmed from the alignment using trimAl v1.5 [49]. To account for potential heterogeneity in evolutionary rates across AUREO domains (bZIP, linker sequence, LOV), we defined alignment-based domain partitions for subsequent phylogenetic inference. Partitioned maximum likelihood phylogenetic inference was performed in IQ-TREE2 v2.3.6 using ModelFinder with partition merging enabled, with branch support estimated from 1,000 ultrafast bootstrap replicates and 1,000 SH-aLRT replicates [50]. In congruence with the species tree, the AUREO tree was rooted using the *Platysulcus tardus* AUREO sequence. Gene duplication histories were reconstructed with DLCpar v2.0.1 in order to obtain primary annotations and a locus-specific reconciled gene tree [51]. Retention scores were then calculated for each duplication event, representing the proportion of descendant species that retained paralogs in both post-duplication lineages. To evaluate the robustness of each duplication event with respect to phylogenetic uncertainty, we additionally reconciled every IQ-TREE-derived ultrafast bootstrap gene tree using Treerecs v1.2 [52], and assessed the robustness of each duplication event by determining how often an equivalent duplication was recovered across the reconciled bootstrap trees. This provided a direct estimate of duplication support under phylogenetic uncertainty. Environmental AUREO sequences of likely heterotrophic stramenopile origin were incorporated by adding them to the trimmed AUREO reference alignment with MAFFT v7.490 with the –add and –keeplength options, preserving alignment length and column structure. The added sequences were then placed on the prior established AUREO reference tree topology with IQ-TREE2 v2.3.6 (see above). The AUREO reference tree was thereby used as a fixed backbone, and the same domain partitions, model-selection settings, support metrics and outgroup were applied as described above. All trees and duplication annotations were visualised either with FigTree v1.4.4 (https://github.com/rambaut/figtree) or with R v4.4.3 using the treeio, ggtree, ggplot2 and ape packages [53–56].

Domain specific phylogenies of LOV-domain containing proteins including AUREOs

In order to identify fine-scale motif variation that could reveal functional differences and local adaptations to polar conditions within the LOV and bZIP domains, we constructed trees that represent sequence space clusters with Profileview [57]. We used this approach once for all LOV-domains with

taxonomic annotations retrieved from the metatranscriptomes but excluded AUREOs due to the high number of sequences. Instead, we added reference AUREOs from *Phaetodactylum tricornutum* (Joint Genome Protein Identifiers 49116, 49458, 56742, and 56060) [58]. This served for identifying the similarity of the blue-light sensing motif. In a second approach, we included AUREOs from our sequencing approaches, i.e. representing Southern Ocean species and specifically closely related non-polar species with further representative AUREO sequences of distinct stramenopile lineages. This approach served to identify likely local adaptations to polar light regimes in AUREOs as well as to resolve a general functional clustering of the AUREO isoforms across different lineages. Therefore, AUREO sequences were retrieved from transcriptomic data of public databases (EukProt) and data from own sequencing approaches were filtered for the blue-light sensing motif with hidden Markov models [30]. For the LOV and bZIP-domains, separate trees were constructed using Profileview [57]. To obtain Profileview-trees of the LOV and the bZIP-domain of sub selected AUREOs, two domain model library were built using Pfam domain libraries for the PAS (PF00989.29), PAS_2 (PF08446.15), PAS_3 (PF08447.16), PAS_4 (PF08448.14), PAS_5 (PF07310.17), PAS_6 (PF08348.15), PAS_7 (PF12860.11), PAS_8 (PF13188.11), PAS_9 (PF13426.11), PAS10 (PF13596.10), PAS_11 (PF14598.10), PAS_12 (PF18095.5), GdpP_PAS (PF21370.1), PdeA_PAS (PF21815.1), and bZIP_1 (PF00170.25), bZIP_2 (PF07716.19), bZIP_C (PF12498.12), and bZIP_Maf (PF03131.21) domain, respectively. For sequences containing more than one LOV-domain (metatranscriptome data), single LOV-domains were extracted from the hmm-search result with additional 25 amino acids as flanking regions. All trees were visualised in R v4.4.1. using the ggtree, ggtreeExtra, ggnewscale, ggplot2, tidyverse, and ape packages [53–55,59–61].

Results

The LOV-domain richness is higher in polar regions

Metatranscriptomic data from the “Sea of Change” project were analyzed with respect to the protein diversity of LOV-domain containing proteins, as well as the respective domain architecture (Supplementary Data 1). In particular, the analysis aimed at discovering novel domain combinations of LOV-domains with other domains, and the distribution of LOV-domain containing protein diversity from pole-to-pole. This approach aimed at providing insight into possible adaptation to polar light conditions (Fig. 1). Sequence samples from the Arctic, Antarctic, and non-polar regions were analyzed in this regard, and then divided according to their origin above or between the polar circles (66.565° for the Arctic and 66.33° for the Antarctic). The area between the northern and the southern polar circles represents the non-polar region (Supplementary Fig. 1). We observed a higher diversity of contigs encoding proteins with LOV-domains in samples from the polar oceans (Fig. 1A), indicating a possibly more important role of blue-light sensing elements in the polar regions that are characterized by lower solar elevation angles and polar light conditions. We generally detected a large variety of LOV-domain proteins with different additional functional domains. The combination of a C-terminal bZIP-domain in combination with a N-terminal LOV-domain, indicative for AUREOs, was found to be the most frequent LOV-domain

combination, followed by LOV-domains either linked to a Histidine Kinase domain, a HATPase_c domain, a WRKY domain, or a HSF-DNA-binding domain (Fig. 1B).

To further classify the LOV-domains of the detected proteins, we constructed a tree that represents sequence space clusters using Profileview [57] (Fig. 1C). Only eukaryotic sequences that could be further classified at the phylum level (based on the MMseq2 classification) were included in the dataset (Fig. 1C). In case of proteins with more than one LOV-domain, e.g. phototropins, the respective LOV-domains were analyzed separately before constructing the sequence space tree. The obtained tree distinguished between LOV-domain proteins combined with diverse protein kinases and those combined with DNA binding domains (Fig. 1C). However, the LOV-domains of the AUREOs clustered with the LOV-protein kinases rather than with other DNA-binding domain proteins.

AUREO duplication history and distribution across the stramenopiles

Stramenopiles belong to the SAR clade (Stramenopiles, Alveolates, Rhizarians). Within the Stramenopiles, two major groups exist which are the Girysta and Bigyra. The exact phylogeny of all groups within the Girysta and Bigyra is under constant change, since new species are still discovered and change the order of phylogenetic trees. The Bigyra group only harbors organisms which are heterotrophic while the Gyrista consist of Pseudofungi (Oomycetes and Hyphochytriomycetes) and the photosynthetically active Ochrophytes [10]. The recently discovered heterotroph stramenopile *Platysulcus tardus* is thought to be a deep-branching stramenopile, branching outside the Bigyra and Gyrista [10,62].

In total, we screened 201 stramenopile protein datasets (*P. tardus*, 30 Bigyra, and 170 Girysta), and retained 96 species (*P. tardus*, 5 Bigyra and 91 Girysta) with a total of 323 AUREO sequences (*P. tardus*, 7 Bigyra and 315 Gyrista). Unexpectedly, two Bigyra lineages (Nanomonadea and Bikosia), which are entirely heterotrophic and lack plastids, apparently possess genes encoding AUREOs. Also, the heterotrophic, deep-branching last common ancestor of the stramenopiles, *P. tardus*, possesses one gene encoding a potential AUREO protein (Fig. 2). Within the Bigyra, AUREOs seem to be absent in the Placididea, Opalinata (subgroups of the Opalozoa, Bigyra) and all Sagenista (Bigyra) lineages (Fig. 2A). Within the Gyrista, we could detect AUREOs in all major photosynthetic lineages (i.e. the Ochrophyta), while no genes were found that could be assigned to the non-photosynthetic Oomycetes and Hyphochytriomycetes (Fig. 2A). Our maximum-likelihood phylogeny of all identified AUREO sequences places almost all of them (313 out of 323) within the clades corresponding to the well described AUREOs 2, 1a, 1b, and 1c isoforms. The 10 early branching AUREOs belong to *P. tardus*, the Bigyra lineage genera MAST3C, MAST3A, *Halocafeteria*, and *Cafeteria* and the Gyrista genera *Octactis* (Dictyochophyceae), *Chrysosystis* and *Aureoumbra* (both Pelagophyceae). The combined results of the gene duplication analysis (DLCpar, Treerecs) identified three non-terminal duplication events (i.e. with at least five descendant taxa) across the AUREO gene family, with a re-identification greater than 85% across bootstrapped tree topologies and with a retention of the descendant gene duplicates greater than 20% across descendant taxa. Three of these duplication events occurred before the functional diversification of the AUREOs (see Fig. 2B, red dots), while one occurred within the AUREO 2–Chrysista

clade. The duplication events that preceded the functional diversification of the AUREOs indicates that AUREO 2 is the earliest-branching AUREO, separating it from the AUREO 1 clade, with a retention of descendant taxa of 74%. Another duplication event at the base of the AUREO 1 clade led to the separation of AUREO 1a and AUREOs 1b and 1c, with a retention across descendant taxa of 73%. A final duplication event then separates AUREO 1b and 1c. However, both genes of this duplication event remained in 23% across descendant taxa, hence the retention is therefore much lower than that of the AUREO 2 and AUREO 1a split. Despite possible deviations resulting from missing transcripts, we overall detected the four AUREOs in a similar total number of species (see Supplementary Data 2), indicating that the four types of AUREOs can be found in most photosynthetic stramenopiles.

Phylogenetic relation of AUREOs based on the LOV-domain

In order to compare the LOV-domain of polar and non-polar AUREO sequences more closely, a Profileview-tree was created based exclusively on the LOV-domain of the AUREOs (Supplementary Data 3) [57]. This approach served to reveal whether the polar AUREO sequences separate from the non-polar ones in terms of the LOV-domain and the associated adaptation to polar light conditions. The AUREO sequences used for the LOV-domain analysis were categorized into the four isotypes based on full-length sequences. The results show that the LOV-domains of AUREO 1a and AUREO 1c are more similar to each other than to the other AUREOs (Fig. 3). LOV-domains of AUREO 1b and 2 can be found in several clades of the tree. The AUREO 2 LOV-domains of centric and pennate diatoms are clearly separated, but there is no separation between LOV-domains from polar or non-polar species (Fig. 3). The higher variability of the LOV-domains of AUREO 2 in pennate and centric diatoms could indicate a functional difference of the LOV-domain. The LOV-domain of *P. tricornutum* AUREO2 is unable to bind FMN and therefore cannot perceive blue light [8]. Although this finding is limited to *P. tricornutum* and *Vaucheria frigida*, it is nevertheless possible that there is a lower selection pressure on the AUREO2-LOV domain, increasing the likelihood that this domain reflects more the evolution of the respective species.

The bZIP-domain is more variable within the AUREO isoforms than the LOV-domain

In a next step, we calculated a further Profileview-tree based on the AUREO bZIP-domains in order to prove whether LOV- and bZIP-domains in the AUREOs show a co-linear evolutionary development [57]. This analysis includes the same sequence data set as used for the LOV-domain analysis. We found that the bZIP-domains of AUREO 1c and AUREO 2 are more conserved than those of the other AUREOs and that they are located mostly within the same clade (Fig. 4). The bZIP-domain of AUREOs 1a and 1b, instead, are both located in separate clades, indicating a more variable bZIP-domain that may be related to the requirements of specific promoter DNA binding sites. We could not detect any differentiation between the bZIP-domains of polar and non-polar AUREOs. Thus, AUREO bZIP-domains appear to be conserved across polar and non-polar environments. The bZIP-domains of almost all heterotrophic Stramenopile sequences cluster together, indicating that similar regulative motifs within the promoters are used.

Discussion

Life in the polar regions is challenged by low temperatures and the radically changing light conditions in the winter and summer seasons with permanent light and absolute darkness, respectively. Another obstacle is the light conditions with a low solar angle requiring a longer passage of photons through the water column, increasing the share of blue light compared to other wavelengths, likely increasing their significance as environmental cue [63]. By integrating global pole-to-pole metatranscriptome with phylogenetic and protein domain-level analysis we show that high-latitude light conditions are associated with greater LOV-domain protein diversity, while AUREOs form a globally dominant and conserved blue-light receptor family expanding towards early-branching heterotroph stramenopiles.

1. Expansion of blue-light sensing proteins towards higher latitudes

Our global pole-to-pole metatranscriptome analysis reveals two key patterns. Firstly, the diversity of LOV-domain-containing proteins apparently increases towards high latitudes (Fig. 1A), suggesting that ecological adaptations involving blue-light photoreception could become more important under polar light regimes. Secondly, the AUREOs emerged as the most frequently detected LOV-domain receptor family across all stations (Supplementary Fig. 3), indicating that they play a key role in light sensing in marine stramenopiles.

The higher richness of LOV-domain proteins at polar latitudes aligns with a general stronger selection for short-wavelength (blue) light-responsive regulation under polar light regimes. Indeed, besides radically changing light conditions in the winter and summer seasons, polar light conditions with a low solar angle require a longer passage of photons through the water column, increasing the share of blue light compared to other wavelengths [64]. The underwater spectral light field also changes depending on ice cover and ice melt, requiring distinct physiological adaptations in polar phytoplankton [65,66]. Also during the polar winter, lunar illumination has been shown to provide sufficient and deep-penetrating blue light to act as an environmental cue for meso- and macrozooplankton [67,68]. Recent investigations have shown that the diatom *Fragilariopsis cylindrus* can switch to a hypometabolism during polar winter regimes. In this state, the diatom cells enter a state of quiescence with reduced metabolic and transcriptional activity, during which no cell division occurs [69]. In sum, the significance of adapting to short-wavelength and to the extreme seasonality of polar light regimes could have been a major driver of LOV-protein diversification, accounting for the observed increase in receptor architecture diversity toward the poles.

Based on the detection frequency across all stations, we found AUREO proteins with their characteristic combination of N-terminal bZIP-domain together with a C-terminal LOV-domain, to emerge as the most frequently detected LOV-receptor proteins from pole-to-pole (Fig. 1B, Supplementary Fig. 3). Based on the taxonomic transcript annotation, Stramenopiles, together with Alveolates and Haptophytes are dominating the dataset (Supplementary Fig. 2), possibly explaining the widespread detection of AUREOs. Yet the dominant prevalence of AUREOs relative to other LOV blue-light sensing proteins suggests a conserved and central position in stramenopile photobiology. We are aware that the detection frequencies reflect the transcript presence and not the genomic presence. Hence, the frequencies might be influenced by the temporal and physiological state of the community. The most frequently found

domain combinations besides AUREOs include a LOV-Histidine-Kinase-HATPase-C-Response-Regulator, WRKY-LOV, and LOV-HSF-DNA-binding (Fig. 1b). These combinations confirm the findings of Coesel et al. [20]. We could also confirm the existence of phototropins (LOV-LOV-pKinase), helmchromes (RGS-LOV-LOV), EF-hand-LOV, Zeitlupe (LOV-Fbox), LOV-Neuralized-LOV, and Homeodomain-LOV motifs in the unicellular protists species of the metatranscriptome datasets, comparable to what is described by Coesel et al. [20]. We further detected completely novel domain combinations, like LOV-CCT (CONSTANS, CO-like, and TOC1), which could be related to the circadian clock of algae [70] (Supplementary Fig. 3), C2 + LOV (PF00168, a Ca^{2+}-dependant membrane-targeting phospholipid binding domain), dimethylbenzinoamidozole phosphoribosyltransferase (DBI_PRT, PF02277) + LOV, DHBP synthase (PF00926) + LOV, polyketide cyclase (PF03364) + LOV, and SnoaL (PF07366) + LOV, also a polyketide cyclase (Supplementary Fig. 3). Most of these proteins were found only once or a few times in the metatranscriptome dataset, thus representing rather lowly expressed transcripts, and we cannot exclude completely the possibility of gene individual assembly errors. Interestingly, a domain combination of an C-terminal LOV-domain (PF13426) and a N-terminal Polyketide cyclase (PF03364) was found 13 times in the metatranscriptome dataset. This combination is also found in a predicted protein from *Noctiluca scintillans* (accession number A0A7S1FCW1, gene NSCI0253_LOCUS31986, retrieved from InterPro). The specific domain architecture, to our knowledge, has not been functionally characterized or discussed in literature to date. The dinoflagellate *Noctiluca scintillans* produces a blue bioluminescence with a peak around 470–480 nm [71] and thus, this domain combination could induce bioluminescence-dependent secondary metabolite synthesis. The domain combination of an N-terminal DHBP-synthase (PF00926) with a C-terminal LOV-domain (PF13426) was found three times in the metatranscriptome dataset. This previously unknown fusion protein could be involved in secondary metabolite production [72], for example in optimizing riboflavin production in response to the daylight cycle. Moreover, a combination of a DHBP synthase followed by a LOV-domain and a GTP cyclohydrolase II (PF00925) could be found in the stramenopile Thraustochytrida. This domain combination also implies a candidate blue-light-regulated riboflavin synthesis with a potential feedback loop, since riboflavin can, besides FMN or FAD, also bind noncovalently to the LOV-domain.

Taxonomically determined LOV-domain sequences from the metatranscriptomic data were analyzed comparatively based on their LOV-domain amino acid sequences. Thereby, the PtAUREO LOV-domains are clustered most closely with the Chlorophyta LOV-domain proteins rather than the LOV-domain of DNA-binding-elements of Haptophyta (Fig. 1C). These LOV-domains belong to the phototropin LOV2-domains, as phototropins contain a LOV1 and LOV2 domain [73]. A reason for this could be the shared functional constraints for blue-light sensing coupled to transcriptional regulation in PtAUREOs and Chlorophyta LOV-domain proteins, which could drive a rather convergent evolution of the LOV-domain despite taxonomic distance, especially since the LOV2-domain is the primary blue-light sensing domain of phototropins.

2. Early diversification and long-term retention of AUREOs in stramenopiles

Consistent with this interpretation, our phylogenetic reconstruction indicates an early origin and a broad distribution of AUREOs across lineages, with deep retention and duplication-associated diversification, alongside a strong sequence conservation of the key domains.

The phylogenetic analysis of the full-length AUREOs reveals deep, well supported clade-level duplications with each major clade containing orthologs from diverse algal lineages, indicating a deep evolutionary origin of AUREOs. The first observed gene duplication separates AUREO1 from AUREO2 proteins, which then separate further by duplication events into the subclades of AUREO 1a and AUREO 1b / AUREO 1c. In addition to these three major duplication events, AUREOs are characterized by further lineage-specific expansions and gene losses within the AUREO clades. Such within-lineage duplications are more recent, occurred after the divergence of the respective taxa, and likely expand blue-light signaling repertoires beyond the conserved AUREO1a/1b/1c/2 duplications. For example, in the Xanthophyceae *Vaucheria frigida*, only two AUREOs have been described (VfAUREO1, VfAUREO2) [5] likely resulting from an early separation of AUREO1 and AUREO2, grouping VfAUREOs into pre-duplication AUREOs. However, the *Vaucheria litorea* transcriptome features 4 AUREO sequences (AUREO2, a duplicated AUREO1a and AUREO1c, see Supplementary Data 2). Other lineages, like the Bacillariophyceae *Fistulifera solaris*, genes encoding 8 AUREOs (2 of each AUREO1a/1b/1c/2) have been detected in the genome, whereas the related benthic diatom *Seminavis robusta* only harbors 3 genes encoding AUREOs, having apparently lost the gene for AUREO 1c (see Supplementary Data 2). In general, gene duplications play a key role in driving functional diversification and specialization [74]. The high retention rates of the duplicated AUREO-genes suggest strong selective advantages, which supports the idea that they function as stable regulatory components within stramenopiles rather than short-lived or transient adaptations. The different topology and AUREO grouping of our phylogeny compared to that of Wu et al. [75] is most likely due to the larger sequence and taxonomic coverage (323 sequences compared to 44 in Wu et al. [75]), and the inclusion of a deep and early branching stramenopile (*P. tardus*) in our phylogeny and duplication analysis.

Notably, early-diverging AUREOs were also found in primary heterotrophic stramenopiles, including *P. tardus* and in non-photosynthetic taxa within the Opalozoa (Fig. 2). We also retrieved early-diverging AUREO sequences from the metatranscriptomes that cluster within the MAST3 lineages (subphylum Opalozoa), underscoring that these heterotrophic AUREO genes are expressed in environmental contexts and serve an ecological function (Supplementary Fig. 5).

The finding of early-diverging AUREO transcripts in the non-photosynthetic stramenopiles is opening a new view on the evolution of AUREOs. As *P. tardus* is a basic lineage of stramenopiles [10,62], this suggests either a retention of an ancestral regulatory gene at the base of the stramenopiles, or that it is a remnant inherited from a photosynthetic ancestor. The latter phenomenon is described for other non-photosynthetic lineages with reduced plastids like dinotoms [76]. In line, our phylogenetic analysis challenges the current assumption that AUREOs are restricted to phototrophs [6]. While stramenopiles most likely originate from a secondary endosymbiosis involving the uptake of a red alga by a eukaryotic

cell and the conversion into a plastid [77], it is not resolved yet if there is an ancestral AUREO protein in red algae, and whether its appearance correlates with the new photosynthetic abilities after secondary endosymbiosis [5,78]. While our metatranscriptome analysis did identify the presence of a transcript combining a LOV-domain with a transcription factor domain in red algae (Fig. 1C), the transcription factor domains are different (zf_H2C2_5 (PF13909) vs bZIP). Further, the red algal LOV-domain rather clusters with haptophyte-derived LOV/DNA-binding domains, whereas the AUREO LOV-domains cluster with chlorophyte-derived LOV-domains. The presence of AUREOs in non-photosynthetic stramenopiles, and their distinction from red-algal blue-light sensing proteins suggests that AUREOs likely evolved within the common ancestor of phototrophic and heterotrophic stramenopiles. This further broadens the functional role and ecological significance of AUREOs which likely act as a general light-sensing signal transducer. For heterotrophic stramenopiles, AUREOs might encompass light sensing, co-regulation, circadian control, as well as behavioral or metabolic regulation, functions well described for e.g. fungi [79] and also heterotrophic protists like *Oxyrrhis marina* in which rhodopsins regulate phototaxis [80].

3. Molecular constraints and regulatory flexibility in AUREO domains

Phylogenetic analyses of the 128 amino acid-long LOV-domains reveal that the AUREO1a isotype is the most conserved, i.e. forming a uniform cluster, regardless of the sampling site. It is followed by the relatively stable conservation of AUREO1c, while AUREO1b and AUREO2 exhibit greater sequence variability within their isotypes (Fig. 3). Differences in AUREO2 LOV-domains correlate with diatom morphology and phylogeny, distinguishing pennate from centric diatoms, yet no clustering by origin is observed for any isotype.

AUREO 1a and 1c, likely experienced strong conserved selection on the amino acid sequence level to preserve precise blue-light sensing functions: studies in *P. tricornutum* link them to light-independent circadian rhythms, with PtAUREO1a influencing other AUREO expression patterns, photosynthetic acclimation, cell cycle regulation, and dimerization with transcription factors [7,8,11,81]. In contrast, the constitutively expressed AUREO2 in *P. tricornutum* may lack FMN binding for direct photoregulation, reducing selective pressure and permitting higher variability consistent with a stabilizing or modulatory role [8].

Recent analyses demonstrate that bZIPs of AUREOs generally share close structural and functional similarities with other bZIP transcription factors found across plants and animals, exhibiting comparable dimerization properties and DNA-binding affinities to substrate DNA [82]. Compared to the LOV domains, the bZIP-domains display a distinct conservation pattern: AUREO 1a and 1b exhibit lower conservation, potentially reflecting higher flexibility in DNA interactions or diverse partnerships with transcription factors. For example, AUREO1a is proposed to have a strong impact on the overall gene regulation in *P. tricornutum* and therefore have numerous fitting binding sites [7,83], whereas highly conserved bZIP-domains in AUREO 1c and 2 suggest specific DNA-binding roles, narrower transcriptional targets, or dimerization preferences, enabling interactions with particular promoter sites. Highly conserved bZIP-

domains, as in AUREO1c and 2, could therefore restrict the partner specificity while variable bZIP sequences, like in AUREO1a and b, indicate broader interactions with promoter motifs [84].

Our bZIP-domain similarity analysis further revealed that almost all (except one) heterotrophic AUREOs form a distinct cluster, suggesting that they share a conserved regulatory role distinct from phototrophic AUREOs. The clustering of the heterotrophic bZIP domains also further argues against an origin by phototrophic contamination.

4. Implications for polar adaptation

Our analysis of the LOV and bZIP-domains revealed no major amino acid changes associated with polar environments, suggesting that AUREOs remain conserved across latitudinal gradients. This is the case for our global analysis as well as a species-to-species comparison of closely related polar diatom species with a non-polar species, showing no latitudinal effect. The persistence of the same AUREO isotypes from non-polar to polar regions indicates that adaptation to extreme light regimes is not achieved through sequence-level divergence. Instead, our findings support the idea that regulatory tuning, like differential expression patterns, diel oscillations, post-translational modifications, or altered dimerization dynamics, could mediate functional adaptations in polar light regimes. Hence, the conservation of AUREOs across the latitudes suggests that ecological success of these organisms relies rather on flexible regulatory mechanisms than structural diversification of AUREO proteins.

Conclusion

Overall, our pole-to-pole metatranscriptome analysis reveals a latitudinal increase in the richness of LOV-domain proteins, representing candidate blue-light photoreceptors and suggesting that spectral light sensing by marine protists becomes increasingly important in polar regions. AUREO-type photoreceptors emerge as dominant and consistently detectable blue-light sensors in global phytoplankton communities, with canonical isotypes remaining structurally stable across stramenopiles, consistent with strong selection on both LOV and bZIP-domains and the long-term maintenance of their regulatory functions. The detection of AUREO sequences in early diverging heterotrophic stramenopiles challenges the view that AUREOs are restricted to phototrophs and instead suggests an earlier origin preceding modern photosynthetic lineages. Crucially, the absence of major sequence differences between polar and non-polar AUREOs suggests that adaptation to extreme polar light conditions is based more on regulatory mechanisms, with the ecological success of polar phytoplankton resulting from a conserved AUREO architecture combined with dynamic regulatory fine-tuning.

Declarations

Author Contributions

Conceptualization: S.W., P.G.K; Data Curation: S.W., M.M.; Formal Analysis: S.W., M.M.; Funding Acquisition: S.W., P.G.K; Investigation: S.W., P.G.K., M.M; Methodology: S.W., M.M.; Software: S.W., M.M.; Supervision: S.W., P.G.K; Visualization: S.W., M.M.; Writing - original draft: M.M.; Writing - review & editing: S.W., P.G.K, M.M.

Acknowledgements

We thank Scarlett Trimborn, Jasmin Stimpfle, Sarah Lena Eggers and Alexandra Kraberg from the Alfred-Wegener-Institute for providing the Southern Ocean diatom strains used here for transcriptomic analyses. We gratefully acknowledge Stefan Neuhaus and Lars Harms from the AWI Data Science Support for recommending Iso-Seq analysis tools and for their assistance in optimizing and deploying workflows on the high-performance computing cluster. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the priority program SPP 1158 "Antarctic Research with comparative investigations in Arctic ice areas" by the following grants KR1661-21/1 and WO1892/3 – 1 and the Helmholtz research program “Changing Earth, Sustaining our Future” (subtopic 6.2 “Adaptation of marine life: from genes to ecosystems” in topic 6 “Marine and Polar Life”) of the Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Germany.

Competing interests

All authors declare no competing interests.

Data availability

Transcriptome data generated within this study are available in the Supplementary Data 4. Additional publicly available metatranscriptome data that support the findings of this study are available in the Joint Genome Institute with the Proposal ID 532/300780 (doi: 10.46936/10.25585/60000951), as well as genome and transcriptome data from EUKPROT V3, [30,31], https://doi.org/10.6084/m9.figshare.21586065.v1 (2022), from the National Center for Biotechnology Information, accession number PRJNA706094, from the European Nucleotide Archive (ENA) at the European Molecular Biological Laboratory – European Bioinformatics Institute (EMBI_EBI) accession number PRJEB18576, and from Figshare under the project number 10.6084/m9.figshare.c.5008046.

References

1. Häfker, N.S., et al.: Rhythms and Clocks in Marine Organisms. Annu. Rev. Mar. Sci. **15**, 509–538 (2023). https://doi.org/10.1146/annurev-marine-030422-113038
2. Valle, K.C., et al.: System Responses to Equal Doses of Photosynthetically Usable Radiation of Blue, Green, and Red Light in the Marine Diatom *Phaeodactylum tricornutum*. PLOS ONE. **9**, e114211 (2014). https://doi.org/10.1371/journal.pone.0114211

3. Jaubert, M., Bouly, J.-P., d'Alcalà, M.R., Falciatore, A.: Light sensing and responses in marine microalgae. Curr. Opin. Plant. Biol. **37**, 70–77 (2017). https://doi.org/10.1016/j.pbi.2017.03.005
4. Duanmu, D., Rockwell, N.C., Lagarias, J.C.: Algal light sensing and photoacclimation in aquatic environments. Plant. Cell. Environ. **40**, 2558–2570 (2017). https://doi.org/10.1111/pce.12943
5. Takahashi, F., et al.: AUREOCHROME, a photoreceptor required for photomorphogenesis in stramenopiles. *Proc. Natl Acad. Sci. USA* 104, 19625–19630 (2007). https://doi.org/10.1073/pnas.0707692104
6. Kroth, P.G., Wilhelm, C., Kottke, T.: An update on aureochromes: phylogeny–mechanism–function. J. Plant. Physiol. **217**, 20–26 (2017). https://doi.org/10.1016/j.jplph.2017.06.010
7. Mann, M., et al.: The aureochrome photoreceptor PtAUREO1a is a highly effective blue light switch in diatoms. Iscience. **23** (2020). https://doi.org/10.1016/j.isci.2020.101730
8. Banerjee, A., et al.: Allosteric communication between DNA-binding and light-responsive domains of diatom class I aureochromes. Nucleic Acids Res. **44**, 5957–5970 (2016). https://doi.org/10.1093/nar/gkw420
9. Armbrust, E.V.: The life of diatoms in the world's oceans. Nat. **459**, 185–192 (2009). https://doi.org/10.1038/nature08057
10. Jirsová, D., Wideman, J.G.: Integrated overview of stramenopile ecology, taxonomy, and heterotrophic origin. ISME J. **18** (2024). https://doi.org/10.1093/ismejo/wrae150
11. Schellenberger Costa, B., et al.: Blue light is essential for high light acclimation and photoprotection in the diatom *Phaeodactylum tricornutum*. J. Exp. Bot. **64**, 483–493 (2013). https://doi.org/10.1093/jxb/ers340
12. Madhuri, S., Lepetit, B., Fürst, A.H., Kroth, P.G.: A Knockout of the Photoreceptor PtAureo1a Results in Altered Diel Expression of Diatom Clock Components. Plants. **13** (2024). https://doi.org/10.3390/plants13111465
13. Costa, B.S., et al.: Correction: Aureochrome 1a is involved in the photoacclimation of the diatom *Phaeodactylum tricornutum*. PLOS ONE. 8 (2013). https://doi.org/10.1371/annotation/7000208e-7505-4c2d-beed-fc99236bbe9f
14. Rockwell, N.C., Su, Y.S., Lagarias, J.C.: Phytochrome structure and signaling mechanisms. Annu. Rev. Plant. Biol. **57**, 837–858 (2006). https://doi.org/10.1146/annurev.arplant.56.032604.144208
15. Fortunato, A.E., et al.: Diatom Phytochromes Reveal the Existence of Far-Red-Light-Based Sensing in the Ocean. Plant. Cell. **28**, 616–628 (2016). https://doi.org/10.1105/tpc.15.00928
16. König, S., Juhas, M., Jäger, S., Kottke, T., Büchel, C.: The cryptochrome—photolyase protein family in diatoms. J. Plant. Physiol. **217**, 15–19 (2017). https://doi.org/10.1016/j.jplph.2017.06.015
17. DeOliveira, C.C., Crane, B.R.: A structural decryption of cryptochromes. Front. Chem. **12**, 1436322 (2024). https://doi.org/10.3389/fchem.2024.1436322
18. Huang, K., Beck, C.F.: Phototropin is the blue-light receptor that controls multiple steps in the sexual life cycle of the green alga *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. USA* 100, 6269–6274

(2003). https://doi.org/10.1073/pnas.0931459100

19. Pushkarev, A., et al.: A distinct abundant group of microbial rhodopsins discovered using functional metagenomics. Nat. **558**, 595–599 (2018). https://doi.org/10.1038/s41586-018-0225-9
20. Coesel, S.N., et al.: Diel transcriptional oscillations of light-sensitive regulatory elements in open-ocean eukaryotic plankton communities. *Proc. Natl. Acad. Sci. USA* 118, e2011038118 (2021). https://doi.org/10.1073/pnas.2011038118
21. Poliner, E., et al.: Aureochromes maintain polyunsaturated fatty acid content in *Nannochloropsis oceanica*. Plant. Physiol. **189**, 906–921 (2022). https://doi.org/10.1093/plphys/kiac052
22. Saikkonen, K., et al.: Toward an integrated understanding of how extreme polar light regimes, hybridization, and light-sensitive microbes shape global biodiversity. One Earth. **7**, 1529–1541 (2024). https://doi.org/10.1016/j.oneear.2024.08.002
23. Gilbertson, R., Langan, E., Mock, T.: Diatoms and Their Microbiomes in Complex and Changing Polar Oceans. Front. Microbiol. **13** (2022). https://doi.org/10.3389/fmicb.2022.786764
24. Eickhoff, L., Bayer-Giraldi, M., Reicher, N., Rudich, Y., Koop, T.: Ice nucleating properties of the sea ice diatom *Fragilariopsis cylindrus* and its exudates. Biogeosciences. **20**, 1–14 (2023). https://doi.org/10.5194/bg-20-1-2023
25. Glantz, S.T., et al.: Functional and topological diversity of LOV domain photoreceptors. *Proc. Natl. Acad. Sci. USA* 113, E1442-E1451 (2016). https://doi.org/10.1073/pnas.1509428113
26. Yee, E.F., et al.: Signal transduction in light–oxygen–voltage receptors lacking the adduct-forming cysteine residue. Nat. Commun. **6**, 10079 (2015). https://doi.org/10.1038/ncomms10079
27. Herrou, J., Crosson, S.: Function, structure and mechanism of bacterial photosensory LOV proteins. Nat. Rev. Microbiol. **9**, 713–723 (2011). https://doi.org/10.1038/nrmicro2622
28. Smith, W.L., Chanley, M.H.: Culture of marine invertebrate animals. Springer (1975)
29. Wohlrab, S., Selander, E., John, U.: Predator cues reduce intraspecific trait variability in a marine dinoflagellate. BMC Ecol. **17**, 8 (2017). https://doi.org/10.1186/s12898-017-0119-y
30. Richter, D.J., et al.: EukProt: A database of genome-scale predicted proteins across the diversity of eukaryotes. *bioRxiv*, 2020.2006.2030.180687 (2022). https://doi.org/10.1101/2020.06.30.180687
31. Poh, Y.-P.: EukProt v3: A database of genome-scale predicted proteins across the diversity of eukaryotes. figshare, Online (2022). https://figshare.com/articles/dataset/TCS_tar_gz/21586065
32. Guajardo, M., Groussman, R.D., Vault, D.: Transcriptomes from the diatoms *Thalassiosira* and *Minidiscus* from the English Channel and Antarctica. Zenodo, Online (2021). https://daniel-vaulot.fr/files/papers/Guajardo_2021_SciData.pdf
33. Beszteri, S., Thoms, S., Benes, V., Harms, L., Trimborn, S.: The Response of Three Southern Ocean Phytoplankton Species to Ocean Acidification and Light Availability: A Transcriptomic Study. Protist. **169**, 958–975 (2018). https://doi.org/10.1016/j.protis.2018.08.003
34. Labarre, A., et al.: Comparative genomics reveals new functional insights in uncultured MAST species. ISME J. **15**, 1767–1781 (2021). https://doi.org/10.1038/s41396-020-00885-8

35. Bushnell, B., BBMap:: A Fast, Accurate, Splice-Aware Aligner. Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA, USA (2014). https://www.osti.gov/biblio/1241166
36. Grabherr, M.G., et al.: Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. **29**, 644–652 (2011). https://doi.org/10.1038/nbt.1883
37. Pacific, B., Lima: Demultiplex barcoded PacBio samples. Pacific Biosciences, Menlo Park, CA, USA (2017). https://github.com/PacificBiosciences/barcoding
38. Pacific, B., IsoSeq: Scalable De Novo Isoform Discovery from long-read PacBio sequencing data. Pacific Biosciences, Menlo Park, CA, USA (2018). https://github.com/PacificBiosciences/IsoSeq
39. Li, H., et al.: The Sequence Alignment/Map format and SAMtools. Bioinformatics. **25**, 2078–2079 (2009). https://doi.org/10.1093/bioinformatics/btp352
40. Haas, B.J., TransDecoder: Identify candidate coding regions within transcript sequences. TransDecoder project Broad Institute, Cambridge, MA, USA (2013). https://github.com/TransDecoder/TransDecoder
41. Mistry, J., et al.: Pfam: The protein families database in 2021. Nucleic Acids Res. **49**, D412–d419 (2021). https://doi.org/10.1093/nar/gkaa913
42. Mirdita, M., Steinegger, M., Breitwieser, F., Söding, J.: Levy Karin, E. Fast and sensitive taxonomic assignment to metagenomic contigs. Bioinformatics. **37**, 3029–3031 (2021). https://doi.org/10.1093/bioinformatics/btab184
43. Van Vlierberghe, M., Di Franco, A., Philippe, H., Baurain, D.: Decontamination, pooling and dereplication of the 678 samples of the Marine Microbial Eukaryote Transcriptome Sequencing Project. BMC Res. Notes. **14**, 306 (2021). https://doi.org/10.1186/s13104-021-05717-2
44. Wood, S.N.: Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models. J. R Stat. Soc. B. **73**, 3–36 (2010). https://doi.org/10.1111/j.1467-9868.2010.00749.x
45. Breitwieser, F.P., Salzberg, S.L.: Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. Bioinformatics. **36**, 1303–1304 (2019). https://doi.org/10.1093/bioinformatics/btz715
46. Emms, D., Liu, Y., Belcher, L., Holmes, J., Kelly, S.: OrthoFinder: scalable phylogenetic orthology inference for comparative genomics. *bioRxiv* (2025). https://doi.org/10.1101/2025.07.15.664860
47. Cho, A., Lax, G., Keeling, P.J.: Phylogenomic analyses of ochrophytes (stramenopiles) with an emphasis on neglected lineages. Mol. Phylogen Evol. **198**, 108120 (2024). https://doi.org/10.1016/j.ympev.2024.108120
48. Katoh, K., Standley, D.M.: MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. **30**, 772–780 (2013). https://doi.org/10.1093/molbev/mst010
49. Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T.: trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. **25**, 1972–1973 (2009). https://doi.org/10.1093/bioinformatics/btp348

50. Minh, B.Q., et al.: IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. **37**, 1530–1534 (2020). https://doi.org/10.1093/molbev/msaa015
51. Mawhorter, R., Liu, N., Libeskind-Hadas, R., Wu, Y.-C.: Inferring Pareto-optimal reconciliations across multiple event costs under the duplication-loss-coalescence model. BMC Bioinform. **20**, 639 (2019). https://doi.org/10.1186/s12859-019-3206-6
52. Comte, N., et al.: Treerecs: an integrated phylogenetic tool, from sequences to reconciliations. Bioinformatics. **36**, 4822–4824 (2020). https://doi.org/10.1093/bioinformatics/btaa615
53. Paradis, E., Claude, J., Strimmer, K.A.P.E.: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. **20**, 289–290 (2004). https://doi.org/10.1093/bioinformatics/btg412
54. Wickham, H.: ggplot2: Elegant Graphics for Data Analysis. Springer, Cham, Switzerland (2016). https://ggplot2.tidyverse.org
55. Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T.-Y.: ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. **8**, 28–36 (2017). https://doi.org/10.1111/2041-210X.12628
56. Wang, L.-G., et al.: Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. Mol. Biol. Evol. **37**, 599–603 (2019). https://doi.org/10.1093/molbev/msz240
57. Vicedomini, R., Bouly, J.P., Laine, E., Falciatore, A., Carbone, A.: Multiple Profile Models Extract Features from Protein Sequence Data and Resolve Functional Diversity of Very Different Protein Families. Mol. Biol. Evol. **39** (2022). https://doi.org/10.1093/molbev/msac070
58. Bowler, C., et al.: The Phaeodactylum genome reveals the evolutionary history of diatom genomes. Nat. **456**, 239–244 (2008). https://doi.org/10.1038/nature07410
59. Wickham, H., et al.: Welcome to the tidyverse. 4, 1686 (2019). https://doi.org/10.21105/joss.01686
60. Xu, S., et al.: Compact Visualization of Richly Annotated Phylogenetic Data. Mol. Biol. Evol. **38**, 4039–4042 (2021). https://doi.org/10.1093/molbev/msab166 ggtreeExtra
61. Campitelli, E., ggnewscale: Multiple Fill and Colour Scales in 'ggplot2'. Zenodo, Online (2024). https://CRAN.R-project.org/package=ggnewscale
62. Shiratori, T., Nakayama, T., Ishida, K.: -i. A New Deep-branching Stramenopile, *Platysulcus tardus* gen. nov., sp. nov. Protist. **166**, 337–348 (2015). https://doi.org/10.1016/j.protis.2015.05.001
63. Ragni, M., Ribera D'Alcalà, M.: Light as an information carrier underwater. J. Plankton Res. **26**, 433–443 (2004). https://doi.org/10.1093/plankt/fbh044
64. Connan-McGinty, S., et al.: Midnight Sun to Polar Night: A Model of Seasonal Light in the Barents Sea. *J. Adv. Model. Earth Syst.* 14, e2022MS003198 (2022). https://doi.org/10.1029/2022MS003198
65. Soja-Woźniak, M., et al.: Loss of sea ice alters light spectra for aquatic photosynthesis. Nat. Commun. **16**, 4059 (2025). https://doi.org/10.1038/s41467-025-59386-x
66. Hoppe, C.J.M., et al.: Photosynthetic light requirement near the theoretical minimum detected in Arctic microalgae. Nat. Commun. **15**, 7385 (2024). https://doi.org/10.1038/s41467-024-51636-8

67. Shen, A., et al.: Altered underwater light characteristics impact photoreceptive mesozooplankton and macrozooplankton from physiological responses to community dynamics. Ecol. Indic. **178**, 114011 (2025). https://doi.org/10.1016/j.ecolind.2025.114011
68. Last, K.S., Hobbs, L., Berge, J., Brierley, A.S., Cottier, F.: Moonlight Drives Ocean-Scale Mass Vertical Migration of Zooplankton during the Arctic Winter. Curr. Biol. **26**, 244–251 (2016). https://doi.org/10.1016/j.cub.2015.11.038
69. Joli, N., et al.: Hypometabolism to survive the long polar night and subsequent successful return to light in the diatom *Fragilariopsis cylindrus*. New. Phytol. **241**, 2193–2208 (2024). https://doi.org/10.1111/nph.19387
70. Liu, H., Zhou, X., Li, Q., Wang, L., Xing, Y.: CCT domain-containing genes in cereal crops: flowering time and beyond. Theor. Appl. Genet. **133**, 1385–1396 (2020). https://doi.org/10.1007/s00122-020-03554-8
71. Letendre, F., Blackburn, A., Twardowski, M.: Linking peak intensity of mechanically stimulated bioluminescence and cell surface area in dinoflagellates. Biol. Open. **14** (2025). https://doi.org/10.1242/bio.062190
72. Sasso, S., Pohnert, G., Lohr, M., Mittag, M., Hertweck, C.: Microalgae in the postgenomic era: a blooming reservoir for new natural products. FEMS Microbiol. Rev. **36**, 761–785 (2012). https://doi.org/10.1111/j.1574-6976.2011.00304.x
73. Freddolino, L., Dittrich, M., Schulten, K.: Dynamic switching mechanisms in LOV1 and LOV2 domains of plant phototropins. Biophys. J. **91**, 3630–3639 (2006). https://doi.org/10.1529/biophysj.106.088609
74. Birchler, J.A., Yang, H.: The multiple fates of gene duplications: Deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. Plant. Cell. **34**, 2466–2474 (2022). https://doi.org/10.1093/plcell/koac076
75. Wu, Y., et al.: Genome-Wide Identification and Analysis of the Aureochrome Gene Family in *Saccharina japonica* and a Comparative Analysis with Six Other Algae. Plants. **11**, 2088 (2022). https://doi.org/10.3390/plants11162088
76. Hehenberger, E., Imanian, B., Burki, F., Keeling, P.J.: Evidence for the Retention of Two Evolutionary Distinct Plastids in Dinoflagellates with Diatom Endosymbionts. Genome Biol. Evol. **6**, 2321–2334 (2014). https://doi.org/10.1093/gbe/evu182
77. Sibbald, S.J., Archibald, J.M.: Genomic Insights into Plastid Evolution. Genome Biol. Evol. **12**, 978–990 (2020). https://doi.org/10.1093/gbe/evaa096
78. Pietluch, F., Mackiewicz, P., Ludwig, K., Gagat, P.A.: New Model and Dating for the Evolution of Complex Plastids of Red Alga Origin. Genome Biol. Evol. **16** (2024). https://doi.org/10.1093/gbe/evae192
79. Corrochano, L.M., Corrochano-Luque, M., Franco-Cano, A., Gutiérrez, G., Cánovas, D.: Light sensing in fungi. Curr. Biol. **35**, R1134–R1138 (2025). https://doi.org/10.1016/j.cub.2025.10.041

80. Hartz, A.J., Sherr, B.F., Sherr, E.B.: Photoresponse in the heterotrophic marine dinoflagellate *Oxyrrhis marina*. J. Eukaryot. Microbiol. **58**, 171–177 (2011). https://doi.org/10.1111/j.1550-7408.2011.00529.x
81. Manzotti, A., et al.: Circadian regulation of key physiological processes by the RITMO1 clock protein in the marine diatom *Phaeodactylum tricornutum*. New. Phytol. **246**, 1724–1739 (2025). https://doi.org/10.1111/nph.70099
82. Khamaru, M., Bose, D., Deb, A., Mitra, D.: Decoding sequence-structure-function-evolution of basic leucine zippers of aureochromes from heterokont algae. J. Struct. Biol. **218**, 108283 (2026). https://doi.org/10.1016/j.jsb.2025.108283
83. Im, S.H., et al.: Identification of promoter targets by Aureochrome 1a in the diatom *Phaeodactylum tricornutum*. J. Exp. Bot. **75**, 1834–1851 (2023). https://doi.org/10.1093/jxb/erad478
84. Nijhawan, A., Jain, M., Tyagi, A.K., Khurana, J.P.: Genomic Survey and Gene Expression Analysis of the Basic Leucine Zipper Transcription Factor Family in Rice. Plant. Physiol. **146**, 323–324 (2007). https://doi.org/10.1104/pp.107.112821

Figures

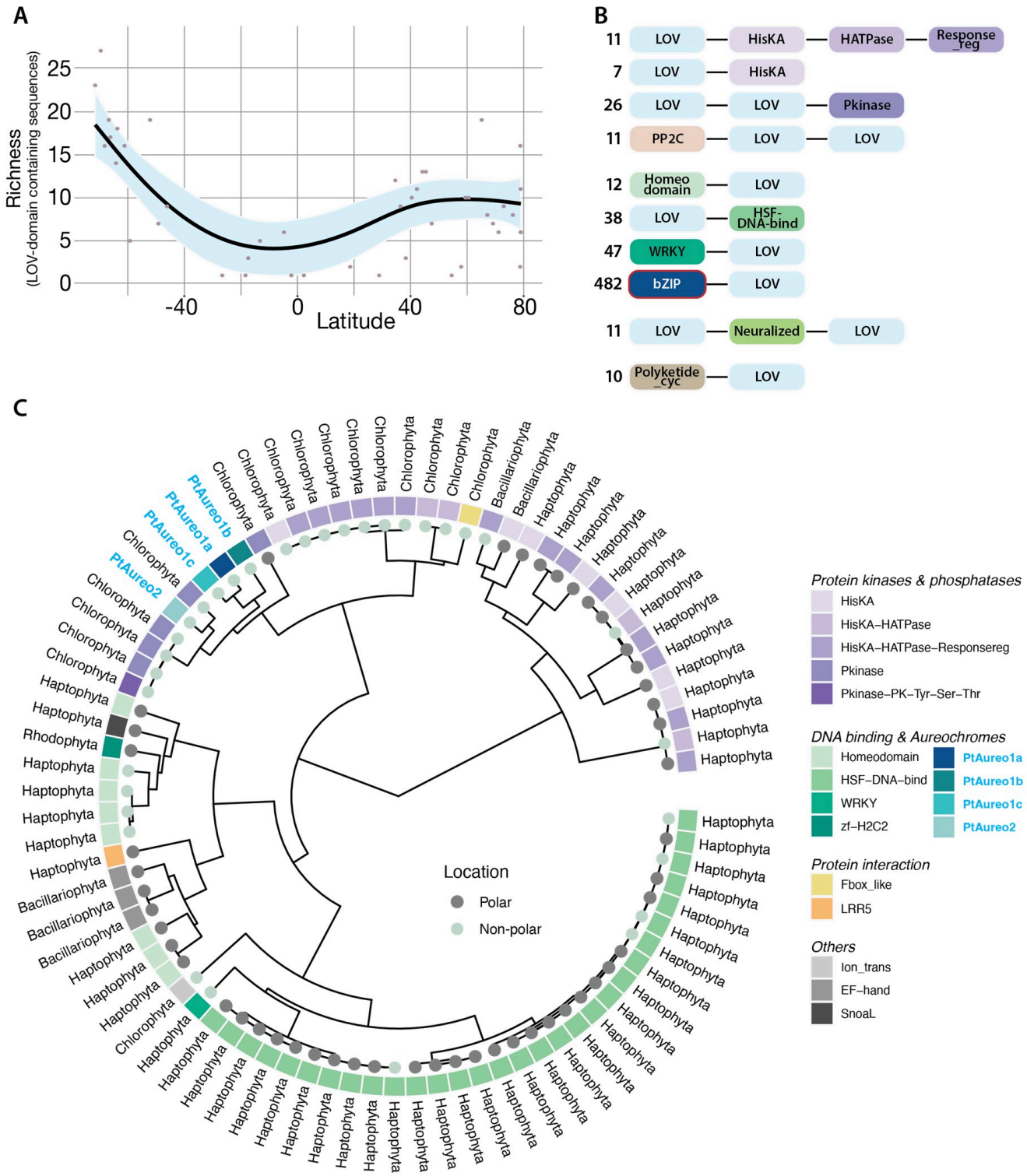


Figure 1

Analysis of metatranscriptomes focusing on transcripts encoding LOV-domains. A) Richness of LOV-domain containing proteins across the latitudes of the globe. B) Overview of the top 10 most frequent found LOV-domain combinations of the whole metatranscriptomic dataset, numbers are detection frequencies. C) Profileview-tree of LOV-domain containing genes that could be taxonomically assigned within the eukaryotes from the pole-to-pole metatranscriptome dataset with *Phaeodactylum tricornutum*

AUREOS as reference (PtAUREOs). Inner circle: Polar or non-polar sequence, Outer circle: Domains found in combination with a LOV-domain.

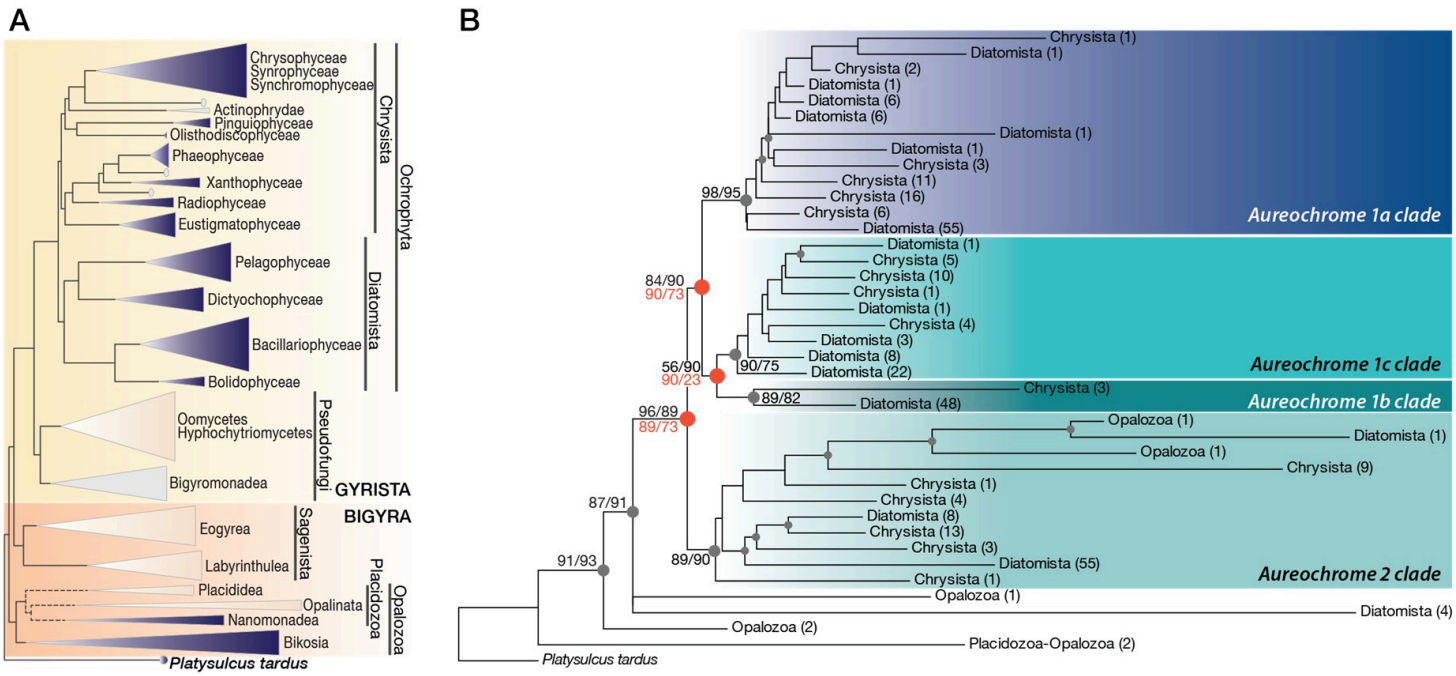


Figure 2

AUREO and early-diverging AUREO sequences within the Stramenopiles. (A) Phylogenetic tree adopted from Cho et al. [47], showing the major taxonomic clades. The dark blue clades contain AUREOs (isoforms 2,1a,1b,1c) and early-diverging AUREO sequences; the light brown clades contain no AUREOs and no reference data were checked for the grey clades. (B) Maximum-likelihood phylogeny of stramenopile AUREO sequences, including duplication retention patterns. Support values (grey, SH-aLTR/UFBoot, see material and methods) are shown for all ancestral internal nodes that define the major, well-described AUREO clades. Red nodes highlight ancestral internal duplication events with high bootstrap reconciliation and retention support (red numbers). Grey dots within the AUREO clades mark nodes with SH-aLTR>70 and UFBoot>85; numbers in brackets correspond to numbers of sequences within the respective clade.

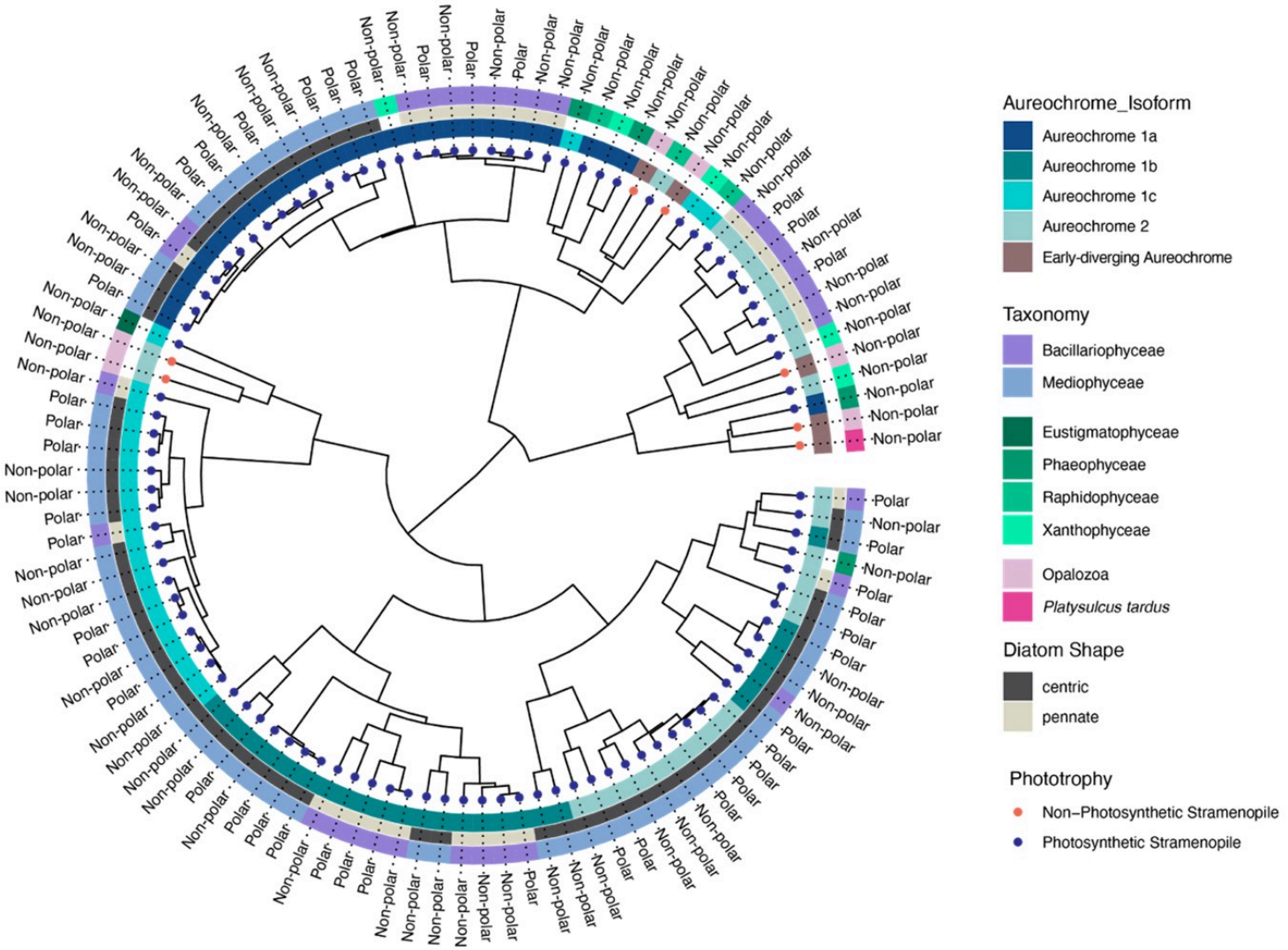


Figure 3

Profileview tree of the LOV-domain of AUREO sequences from transcriptomes and genomes of Southern Ocean and non-polar phytoplankton species. Inner circle: Photosynthetic or Non-photosynthetic Stramenopile, Middle circle: Type of AUREO, Outer circle: Taxonomic group.

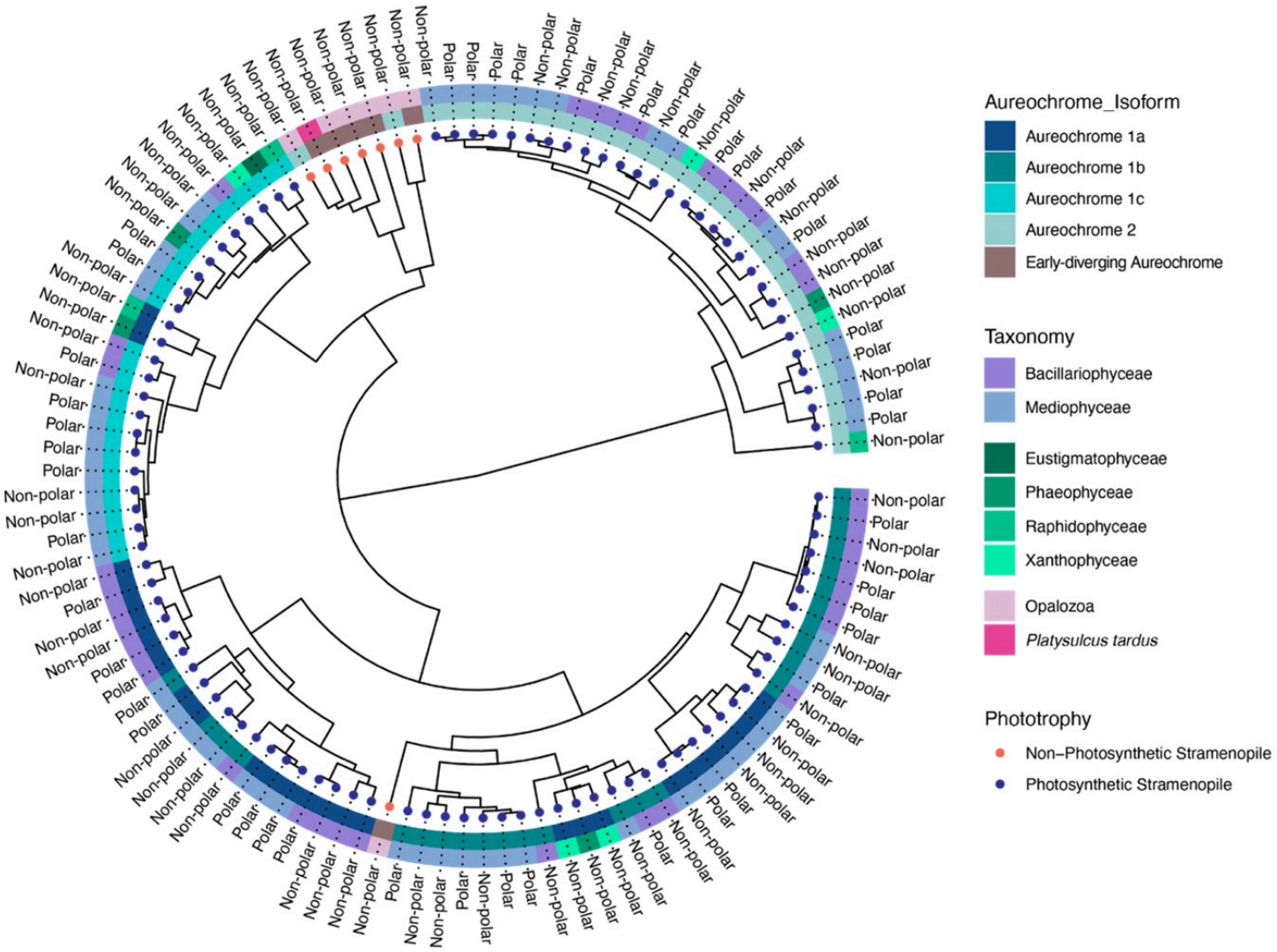


Figure 4

Profileview tree of the bZIP-domain of AUREO sequences from transcriptome and genome phytoplankton species coming from non-polar and Southern Ocean locations. Inner circle: Photosynthetic or Non-photosynthetic Stramenopile, Middle circle: Type of AUREO, Outer circle: Taxonomic group.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryData2.xlsx
- DescriptionofAdditionalSupplementaryFilesSW.pdf
- SupplementaryData4.txt
- SupplementaryInformationSW.pdf
- SupplementaryData1.xlsx

- SupplementaryData3.xlsx