Molecular Plant
Research article

CellPress
Partner Journal

Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity

David R. Nelson[1,2,8,*], Alexandra Mystikou[1,2,3,8,*], Ashish Jaiswal[1], Cecilia Rad-Menendez[4], Michael J. Preston[5], Frederik De Boever[4], Diana C. El Assal[1], Sarah Daakour[1,2], Michael W. Lomas[5], Jean-Claude Twizere[1,6], David H. Green[4], William C. Ratcliff[7] and Kourosh Salehi-Ashtiani[1,2,*]

[1]Division of Science and Math, New York University Abu Dhabi, Abu Dhabi, UAE

[2]Center for Genomics and Systems Biology (CGSB), New York University Abu Dhabi, Abu Dhabi, UAE

[3]Biotechnology Research Center, Technology Innovation Institute, PO Box 9639, Masdar City, Abu Dhabi, UAE

[4]Culture Collection of Algae and Protozoa, Scottish Association for Marine Science, Oban, Scotland, UK

[5]National Center for Marine Algae and Microbiota, Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, USA

[6]Laboratory of Viral Interactomes, GIGA Institute, University of Liege, Liege, Belgium

[7]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA

[8]These authors contributed equally to this article.

***Correspondence: David R. Nelson (drn2@nyu.edu), Alexandra Mystikou (alexandra.mystikou@tii.ae), Kourosh Salehi-Ashtiani (ksa3@nyu.edu)**

https://doi.org/10.1016/j.molp.2024.03.011

ABSTRACT

Macroalgae are multicellular, aquatic autotrophs that play vital roles in global climate maintenance and have diverse applications in biotechnology and eco-engineering, which are directly linked to their multicellularity phenotypes. However, their genomic diversity and the evolutionary mechanisms underlying multicellularity in these organisms remain uncharacterized. In this study, we sequenced 110 macroalgal genomes from diverse climates and phyla, and identified key genomic features that distinguish them from their microalgal relatives. Genes for cell adhesion, extracellular matrix formation, cell polarity, transport, and cell differentiation distinguish macroalgae from microalgae across all three major phyla, constituting conserved and unique gene sets supporting multicellular processes. Adhesome genes show phylum- and climate-specific expansions that may facilitate niche adaptation. Collectively, our study reveals genetic determinants of convergent and divergent evolutionary trajectories that have shaped morphological diversity in macroalgae and provides genome-wide frameworks to understand photosynthetic multicellular evolution in aquatic environments.

Key words: macroalgae, multicellularity, comparative genomics, evolution, adhesome, endogenous viral elements

Nelson D.R., Mystikou A., Jaiswal A., Rad-Menendez C., Preston M.J., De Boever F., El Assal D.C., Daakour S., Lomas M.W., Twizere J.-C., Green D.H., Ratcliff W.C., and Salehi-Ashtiani K. (2024). Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity. Mol. Plant. **17**, 1–25.

INTRODUCTION

Macroalgae encompass a diverse array of organisms belonging to three distinct phyla, each of which has independently evolved multicellularity. These organisms demonstrate extraordinary morphological diversity, ranging from millimeter-scale filaments to colossal 65-m-long kelps, and inhabit a variety of ecological niches. The simpler form of multicellularity, filamentous growth, is common in macroalgae and parallels that seen in other simpler multicellular phototroph lineages, such as Zygnematophyceae (Hess et al., 2022). While many algal lineages have adopted simple multicellularity, complex phototrophic multicellularity is a defining feature of plants and macroalgae (Adl et al., 2012). Shifts to simpler multicellularity from unicellularity (e.g., in *Volvox carterii*) (Prochnik et al., 2010) may not show significant changes in gene content, instead resulting primarily from differential regulation of gene expression (Matt and Umen, 2018) and splicing patterns (Balasubramanian et al., 2023). Indeed, shifts to multicellularity can be observed experimentally with little or no change in basal genetic content (Ratcliff et al., 2012), and some lineages have reverted to unicellular forms after evolutionary periods of multicellularity. Thus, transitions to

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

Please cite this article in press as: Nelson et al., Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity, Molecular Plant (2024), https://doi.org/10.1016/j.molp.2024.03.011

complex multicellularity are more likely to involve newly acquired, lineage-defining gene sets, which can be discovered through extensive genome sequencing.

The three principal phyla (Adl et al., 2019) of marine macroalgae—Rhodophyta (Brawley et al., 2017) (red algae), Chlorophyta (He et al., 2021) (green algae), and Ochrophyta (Bringloe et al., 2020) (brown algae)—exemplify the convergent evolution of complex multicellularity in aquatic phototrophs. Studies on *Volvox carteri*, a simple multicellular green alga, shed light on the early stages of chlorophyte germ-soma differentiation, a key feature of multicellularity (Ferris et al., 2010; Prochnik et al., 2010; Matt and Umen, 2016, 2018; Umen, 2020; Yamamoto et al., 2021; Balasubramanian et al., 2023). The comprehensive analysis of *Volvox* cell-type transcriptomes revealed distinct molecular and metabolic programming between these cell types (Matt and Umen, 2018). Nonetheless, while regulatory changes play a key role in the evolution of multicellularity in the volvocine green algae, no significant changes in gene content were found compared to its closest unicellular relatives. While increased sophistication in ECMs are hallmarks of many simple multicellular lineages, their higher complexity usually derives from gene regulatory changes and not acquisitions (Kloareg et al., 2021).

Macroalgal lineages evolved specific metabolic and morphological traits underlying their multicellular phenotypes, which form independent bases for their evolution of multicellularity. Investigating the genetic foundations of multicellularity in these lineages is crucial for understanding how complex life adapts to varying environmental challenges and for gaining insights into the genetic underpinnings of this significant evolutionary transition. A scarcity of macroalgal genome sequences has thus far constrained the development of a comprehensive comparative analysis of macroalgal multicellularity. Increased genome sample sizes will facilitate the statistical resolution of hypothesis tests regarding phyla-wide gene gains and losses accompanying the rise to multicellularity in the three clades. In this study, we greatly expanded the number of available macroalgal genomes from 14 to 124, shedding new light on the genomic basis of macroalgal multicellularity.

Rhodophyta and Chlorophyta algae belong to the Archaeplastida supergroup descending from the original plastid/nuclear symbiotic event in eukaryotic algae (Yoon et al., 2004; Brawley et al., 2017). In contrast, Ochrophyta contain Rhodophyta algal plastids derived from further endosymbiosis (Bringloe et al., 2020). Ochrophyta evolution through this route may have involved multiple secondary endosymbioses, rather than a single event (Yoon et al., 2004). Macroalgae thrive in various conditions in all climates in marine, freshwater, and brackish environments. They can grow in intertidal and benthic regions as deep as 268 m (Littler et al., 1985). They represent a particularly powerful system for examining the environmental and genetic drivers of multicellular evolution, as independent transitions in these clades were separated by more than a billion years and occurred under fundamentally different environmental conditions (Yoon et al., 2004; Brawley et al., 2017).

The evolution of algal lineages has been meticulously analyzed using molecular clock experiments and geological evidence (Heckman et al., 2001; Douzery et al., 2004; Hedges et al., 2004; Yoon et al., 2004; Berney and Pawlowski, 2006; Zimmer et al., 2007; Herron et al., 2009; Lang et al., 2010; Fiz-Palacios et al., 2011; Gueidan et al., 2011; Parfrey et al., 2011; Blank, 2013; Gaya et al., 2015; Munakata et al., 2016; Yang et al., 2016; Caspermeyer, 2017). Ochrophyta and Rhodophyta diverged ~1.6 billion years ago (bya), and rhodophytes and chlorophytes split ~1.2 bya. Chlorophyta and rhodophyte macroalgae emerged more than 1 bya; in contrast, the stramenopile-origin macroalgal Phaeophyceae only emerged ~200 million years ago (mya) (Supplemental Figure 1). Phaeophyte seaweeds thus evolved in a period with comparably high atmospheric O_2 and solar luminosity but low CO_2 (Supplemental Figure 1). The global prevalence of the recently emerged Phaeophyceae, evident in extensive giant kelp forests and abundant seaweed biomass, indicates they are well adapted to thrive in contemporary environmental conditions. For instance, *Sargassum* emerged only 6.7 mya (Yip et al., 2020), yet has swiftly colonized the world's oceans, generating significant biomass (Arellano-Verdejo and Lazcano-Hernandez, 2021; Liu et al., 2021; Song et al., 2021; Stelling-Wood et al., 2021; Mulders et al., 2022). Invasive phaeophytes threaten various industries and ecosystems; understanding the genomic and evolutionary bases of their success will facilitate management efforts and biotechnological utilization (Charrier et al., 2017).

Our large-scale project, ALG-ALL-CODE-macro, sequenced 110 macroalgal genomes from diverse climates and phyla, offering new insights into macroalgal biology and the evolution of multicellularity. We identified key genes for processes necessary for multicellular phenotypes (e.g., cellular adhesion, transcription factors [TFs]). Many of these had viral origins and were unique to or conserved among the three primary three macroalgal lineages. The diverse viral-origin sequences in various proteins underpin lineage-specific mechanisms that support the evolution of multicellular macroalgae with differentiated, coordinated tissues. Our study illuminates the convergent and divergent evolutionary trajectories shaping macroalgae's diverse morphologies, outlining a basis for understanding photosynthetic multicellular evolution in marine environments.

RESULTS AND DISCUSSION

Comparative genomics of macroalgae

To understand the genomic principles underpinning the evolution of macroalgal multicellularity, we sequenced 110 new macroalgal genomes grown in international culture centers, of which 105 are different species. All available macroalgal cultures from the Culture Collection of Algae and Protozoa (CCAP), Scotland, UK (www.ccap.ac.uk) and the National Center for Marine Algae and Microbiota (NCMA), Maine, USA (ncma.bigelow.org) were cultured for downstream genome sequencing. Five species have been sequenced twice due to their availability and presence in both culture collection centers in separate cultures: *Bostrychia radicans* CCAP1357/2 and CCMP2677; *Catenella nipae* CCMP2850 and CCMP2848; *Porphyra pulchella* CCMP3230 and CCAP1379_3; *Pylaiella littoralis* CCMP1907 and CCAP1330/9; and *Ulvella leptochaete* CCAP6037/1 and CCAP6000/1. One species was sequenced thrice from three different cultures from both culture collection centers: *Bostrychia moritziana* CCAP1357/3,

Please cite this article in press as: Nelson et al., Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity, Molecular Plant (2024), https://doi.org/10.1016/j.molp.2024.03.011

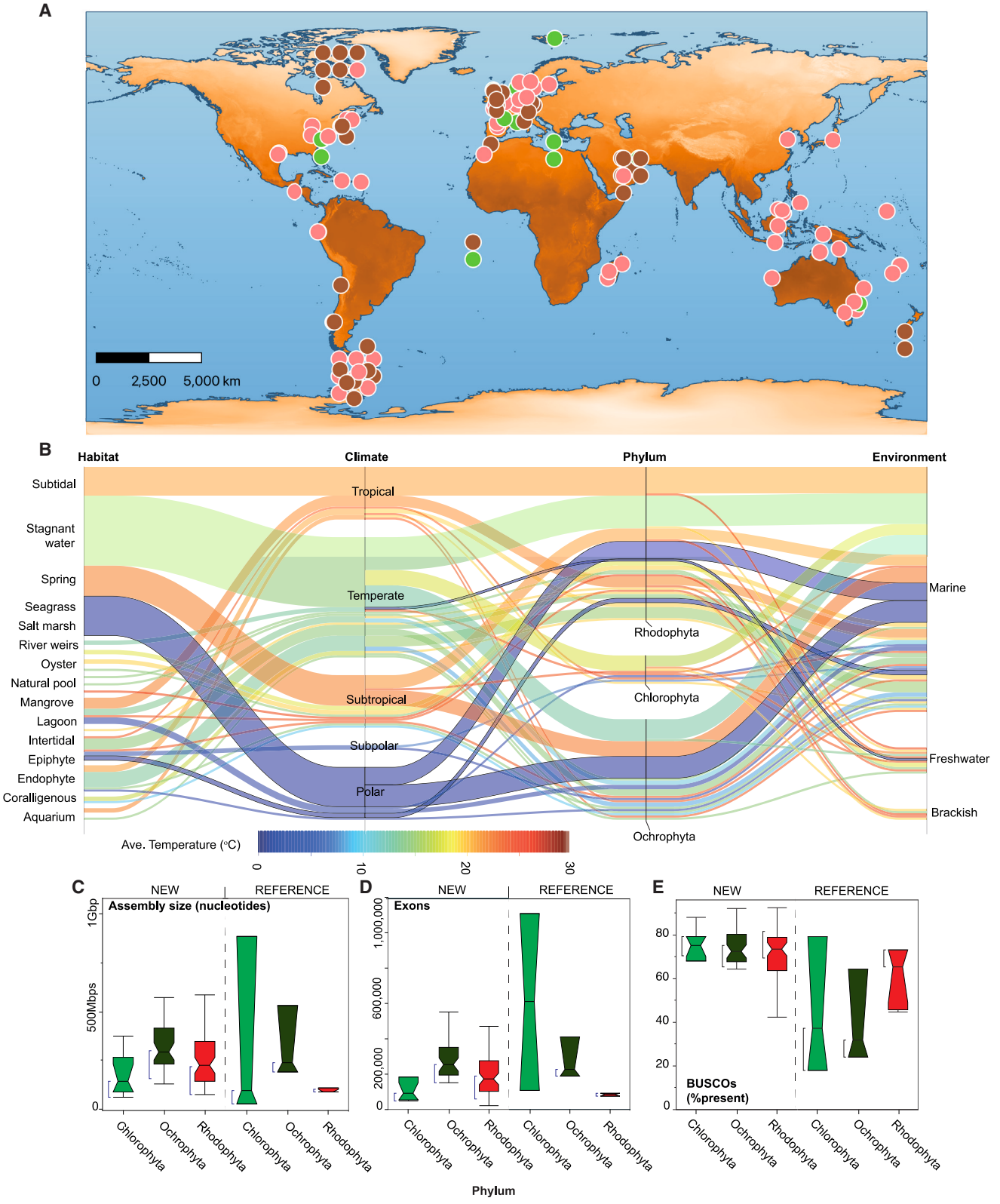


Figure 1. Geographies, habitats, climates, and divergence estimates for phyla of the newly sequenced macroalgae.

(A) Global distribution of the 110 macroalgal isolates whose DNA was extracted to reconstruct genomes for ALG-ALL-CODEmacro by phylum.

(B) Flow diagram showing the native habitat and climate of sequenced macroalgae. The color scale shows the average temperature for all isolates represented by a given band. Distributions of taxa correlate with their availability in culture centers and roughly with their natural distributions, with Rhodophyta having many more extant species than the other clades.

(legend continued on next page)

Please cite this article in press as: Nelson et al., Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity, Molecular Plant (2024), https://doi.org/10.1016/j.molp.2024.03.011

CCAP1357/8 and CCMP3247. The resequencing was done for benchmarking purposes (Supplemental Table 1) and to examine sequences unique to either culture accession. Species from diverse geographies (Figure 1B and 1C, Supplemental Table 1), including those from temperate, tropical, subtropical, polar, and subpolar climates, and from five major latitude parallels were cultured for genome sequencing. These species are from the three major macroalgal phyla that emerged epochs apart in vastly different geological conditions (Supplemental Figure 1).

In addition to the 110 international macroalgal culture samples, nine macroalgal species from field samples we previously recovered from the UAE (https://doi.org/10.5281/zenodo.7758508) were included in our analyses. The expanded genome collection presented here, from 14 to 124 genomes, helps to resolve long-standing questions regarding macroalgal biology and the evolution of multicellularity in different geological and environmental time frames spanning epochs. Our dataset of 110 genomes represents a substantial improvement in both quantity and quality compared to previously sequenced macroalgal genomes (Supplemental Table 1 and Figure 1C–1E). After sequencing, genome assemblies were computationally decontaminated using a pipeline developed in house (Supplemental Figure 2, Supplemental Data 3). This pipeline is available for use in non-macroalgal, eukaryotic species and offers substantial improvements in accuracy and batch processing over other computational decontamination methods. Furthermore, the separated contaminant sequences and their homologs in the National Center for Biotechnology Information (NCBI) non-redundant (nr) protein database are also included in Supplemental Data 3.

As an initial, high-level analysis, we analyzed the degree of shared vs. unique genes by examining the distribution of protein families (PFAMs) in the three lineages using ternary (Ponsen et al., 2009) graphing (Figure 2A) and response screening (Supplemental Table 2). Genes unique to each lineage could play important roles in the independent acquisition of multicellularity in the three lineages, while common PFAMs highlight possible evolutionary convergence (Figure 2). Batch *t-test*s were used to compare PFAMs by phyla. We found that 1065 PFAMs significantly differed between clades (adjusted [adj.] $p < 0.05$), while 9233 PFAMs were significantly equivalent among the three phyla (eq. $p < 0.05$, Supplemental Table 2). These results produced similar findings to the ternary analysis (Figure 2, Supplemental Table 2). Protein families near the center of the ternary graph represent conserved functions, while peripheral PFAMs at the corners of the three lineages indicate lineage-specific inclusions. We extracted PFAM sets at incremental purity cutoffs in each of the three extremes in the ternary graph. Purity is defined by the degree which the PFAM is unique to any given lineage or present at higher copy numbers than the other two. Higher-purity sets represent specialized gene sets for each clade (Figure 2). Sets of filtered PFAMs (>80% purity) from the three lineages were used in hypergeometric distribution tests to analyze functional enrichment in domain-centric Gene Ontology (dcGO) (Fang and Gough, 2013) analyses (Figure 2). The PFAM domains yield information about Gene Ontology (GO) enrichment in a domain-centric solution to functional enrichment analysis. Only the Ochrophyta and Chlorophyta showed enrichment in any GO term at >80% purity. Ochrophyta ($n = 43$ genomes) were enriched for 646 GO terms using the isolated PFAM set as dcGO queries (Figure 2). Ochrophyta-enriched GO terms at a PFAM purity >80% included "multicellular organismal process" (GO: 0032501, $Z = 4.61$, adj. $p = 4.12 \times 10^{-4}$), "multicellular organism development" (GO: 0007275, $Z = 4.18$, adj. $p = 9.35 \times 10^{-4}$), "regulation of multicellular organismal development" (GO: 2000026, $Z = 3.64$, adj. $p = 4.45 \times 10^{-3}$), and "regulation of multicellular organismal process" (GO: 0051239, $Z = 3.36$, adj. $p = 5.98 \times 10^{-3}$).

This relatively large set of GO terms for high-purity PFAMs in Phaeophyceae represents their specialization into specific gene sets facilitating multicellularity compared to the other two macroalgal clades. Lineages with large influxes of new genes, such as the rampant lysogenic virus genome insertions observed in *Ectocarpus*, may have locked themselves into multicellular modes. In contrast, many other lineages have representatives that reverted from multicellularity to unicellularity. The reversion from multicellularity to unicellularity has apparently occurred in cyanobacteria, green algae, and filamentous fungi, while no examples of reversion have been seen in metazoans, land plants, or Phaeophyceae (Peter et al., 2023). Our data provide further evidence for the uniqueness of the genetic basis for multicellularity in the Phaeophyceae.

Chlorophytes were enriched for 28 GO terms at >80% PFAM purity, including "shoot system development" ($Z = 9.76$, adj. $p = 7.96 \times 10^{-6}$), "flower development" ($Z = 8.85$, adj. $p = 1.97 \times 10^{-4}$), "plant organ development" ($Z = 8.21$, adj. $p = 1.31 \times 10^{-4}$), "reproductive shoot system development" ($Z = 8.23$, adj. $p = 3.49 \times 10^{-4}$), "response to jasmonic acid" ($Z = 7.67$, adj. $p = 2.57 \times 10^{-3}$), and "positive regulation of G-protein coupled receptor protein signaling pathway" ($Z = 5.93$, adj. $p = 9.72 \times 10^{-3}$). PFAMs comprising these GO terms are implicated in molecular processes specific to chlorophytes. For example, the GO term for phyllome development (enriched at $Z = 8.27$, adj. $p = 6.17E{-}04$, related to GO:0010432, bract development; GO:0048440, carpel development; GO:0048366, leaf development; GO:0048441, petal development, GO:0048442, sepal development; and GO:0048443, stamen development) is composed of 12 PFAMs and five of these were specific to chlorophyte macroalgae: a TF (PF00847; found in ethylene-inducible TFs) (Ohme-Takagi and Shinshi, 1995), a K-box domain (PF01486; found in SRF-type TFs, e.g., the *Arabidopsis* floral homeotic gene PISTILLATA) (Goto and Meyerowitz, 1994), a B3 DNA-binding domain (PF02362, e.g., ARF1, a TF that binds to auxin response elements) (Ulmasov et al., 1997), and SQUAMOSA-pROMOTER BINDING PROTEINs, which are plant TFs (Klein et al., 1996).

Many processes enriched for in chlorophyte macroalgae were conserved in higher plants, indicating that a propensity to

(C) Genome sizes for the newly sequenced and reference macroalgae. C, Chlorophyta; O, Ochrophyta; R, Rhodophyta.
(D) Exon counts for the newly sequenced and reference macroalgae.
(E) BUSCOs (Manni et al., 2021) (% present = 1 – missing) from the newly sequenced and reference macroalgae.

Please cite this article in press as: Nelson et al., Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity, Molecular Plant (2024), https://doi.org/10.1016/j.molp.2024.03.011

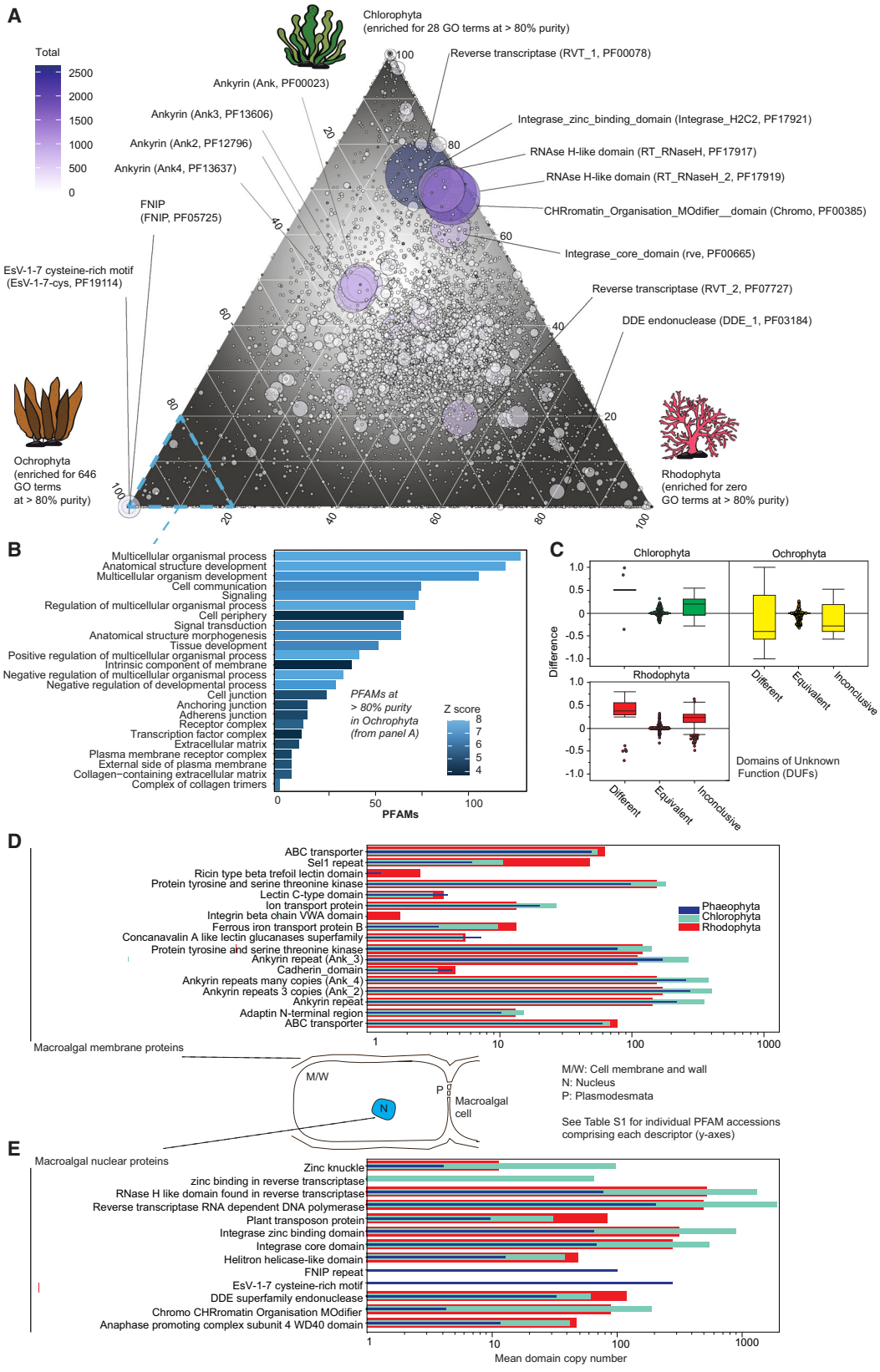


Figure 2. Lineage-specific protein families in macroalgal clades form the basis for their independent paths to multicellularity.

(A) Ternary plot constructed from PFAM count averages for each of Chlorophyta ($n = 11$), Rhodophyta ($n = 74$), and Ochrophyta ($n = 45$). Information about the degree of uniqueness of PFAMs (purity) is shown. Axes indicate the purity of the PFAM for the phyla shown on the triangle points. The plot is

(legend continued on next page)

Please cite this article in press as: Nelson et al., Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity, Molecular Plant (2024), https://doi.org/10.1016/j.molp.2024.03.011

develop these enriched functions may have been latent in the last common ancestor of Chlorophyta seaweeds and land plants (~1.2 bya) (Hedges et al., 2004; Yoon et al., 2004; Zimmer et al., 2007; Parfrey et al., 2011; Yang et al., 2016). Indeed, in some Chlorophyta lineages evolving multicellularity, the base gene set prerequisite appears to have been possessed *in plena* by their unicellular progenitors (Prochnik et al., 2010). Recent phylogenetic reconstructions have suggested that the green macroalgal *Ulva* clades are paraphyletic (Fučíková K, 2014; Fang et al., 2018; Hou et al., 2022); thus, similar phenotypes arose independently from the fertile, proto-multicellular chlorophyte gene repertoire. A recent study found transporters (e.g., nuclear pore proteins), hormone signaling components, and the TALE superclass of homeobox genes were expanded in the green alga *Caulerpa lentillefera* in a manner similar to land plants (Arimoto et al., 2019). These findings concur with our data, showing that many chlorophyte lineages already had a gene set primed for the evolution of multicellularity through plant-like pathways.

Phaeophyte algae's specialization may partly stem from their recent emergence. In comparison, Chlorophyta and rhodophyte multicellularity originated over a billion years ago (Brawley et al., 2017) (Supplemental Figure 1), potentially obscuring informative genomic signals in our ternary analysis, except for shared Chlorophyta/Rhodophyta overrepresented PFAMs compared to Ochrophyta (Fig 2A). These PFAMs include tandem arrays of viral-origin reverse transcriptase (RT), RNAse-like, and integrase domains, often found near centromeres in microalgae and plants (US Patent WO2012061481A2). The rationale for retaining these virus-related elements remains uncertain (Sharma et al., 2013), but our findings indicate they may constitute unique genomic backbones for Archaeplastida macroalgae. Conversely, Ochrophyta and Chlorophyta genomes had the most ankyrin repeats, primarily acquired after Ochrophyta–Rhodophyta and Chlorophyta–Rhodophyta splits. Our prior research showed strong positive selection in marine microalgal ankyrins (Nelson et al., 2021), implying potential fitness advantages for marine macroalgae. Our data suggest that these extensive, lineage-specific duplications of potentially adaptive and often virally sourced genes comprise core genomic features for members of these phyla.

Our data reveal that the recent emergence of multicellular Phaeophyceae is reflected in their genomes through domain-centric functional enrichment (Figure 2B). In other words, the genes uniquely responsible for the Phaeophyceae rise to multicellularity are contained in these modular PFAM sets. Their genomic signal detailing their route to, and maintenance of, a multicellular lifestyle is much newer than rhodophytes and chlorophytes. This signal was seen in the PFAM datasets, wherein a ternary analysis was sufficient to identify Phaeophyceae genes comprising GO terms for multicellular developmental processes. The pervasive sequence insertions containing EsV-1-7 and FNIP motifs in Ochrophyta were not seen in Chlorophyta and Rhodophyta genomes. Phaeophyceae EsV-1-7 proteins included a variety of domains with putative or established roles in multicellular organismal development (Supplemental Table 5). These results further evidence that virus-sourced gene collections underpin the evolution of brown algal multicellularity.

Protein family loss and gain during macroalgal evolution

Comparisons between macroalgal chlorophytes, phaeophytes, and rhodophtyes could yield insight into unique and shared genes in these clades, differentiating them from their microalgal (i.e., microscopic, unicellular) counterparts. Gene circuits regulating the onset of complex multicellular development are conserved in fungal lineages (Nagy et al., 2018); here, we tested for the presence of similar PFAMs found in such circuits that could differentiate macro-from microalgae. Intersection analysis of these PFAM sets revealed their shared and unique retention (Figure 3A). Rhodophyta microalgae had 29 unique PFAMs, while their macroalgal counterparts had 883. Microalgal chlorophytes had 696 unique PFAMs, whereas their macroalgal counterparts had 158 unique PFAMs. The contrasting unique PFAM contents between Rhodophyta and Chlorophyta macro-/microalgae suggest convergence in Chlorophyta seaweeds and functional divergence in Rhodophyta macroalgae, potentially as a function of their evolutionary time spent as multicellular organisms.

Nine PFAMs were unique to microalgae, regardless of lineage (Figure 3A). These included the Parvovirus non-structural protein NS1, the TruB_C tRNA Pseudouridine synthase II, the YojJ bacterial membrane-spanning protein, an alcohol dehydrogenase GroES-associated domain, a SUMO isopeptidase, a Rrp7 RRM-like domain, and two domains of unknown function (DUFs) (DUF2202 and DUF6787). The photosystem II 12-kDa extrinsic protein (PsbU, PF06514) was also lost in all chlorophyte (0% retention rate [RR]) and most ochrophyte 16% RR) macroalgae, but most rhodophyte macroalgae retained PsbU (69% RR; Supplemental Figure 3, Supplemental Data 3). In contrast, PsbP was lost in ochrophyte macroalgae ($p = 5.68 \times 10^{-14}$) but retained in chlorophyte and rhodophyte macroalgae.

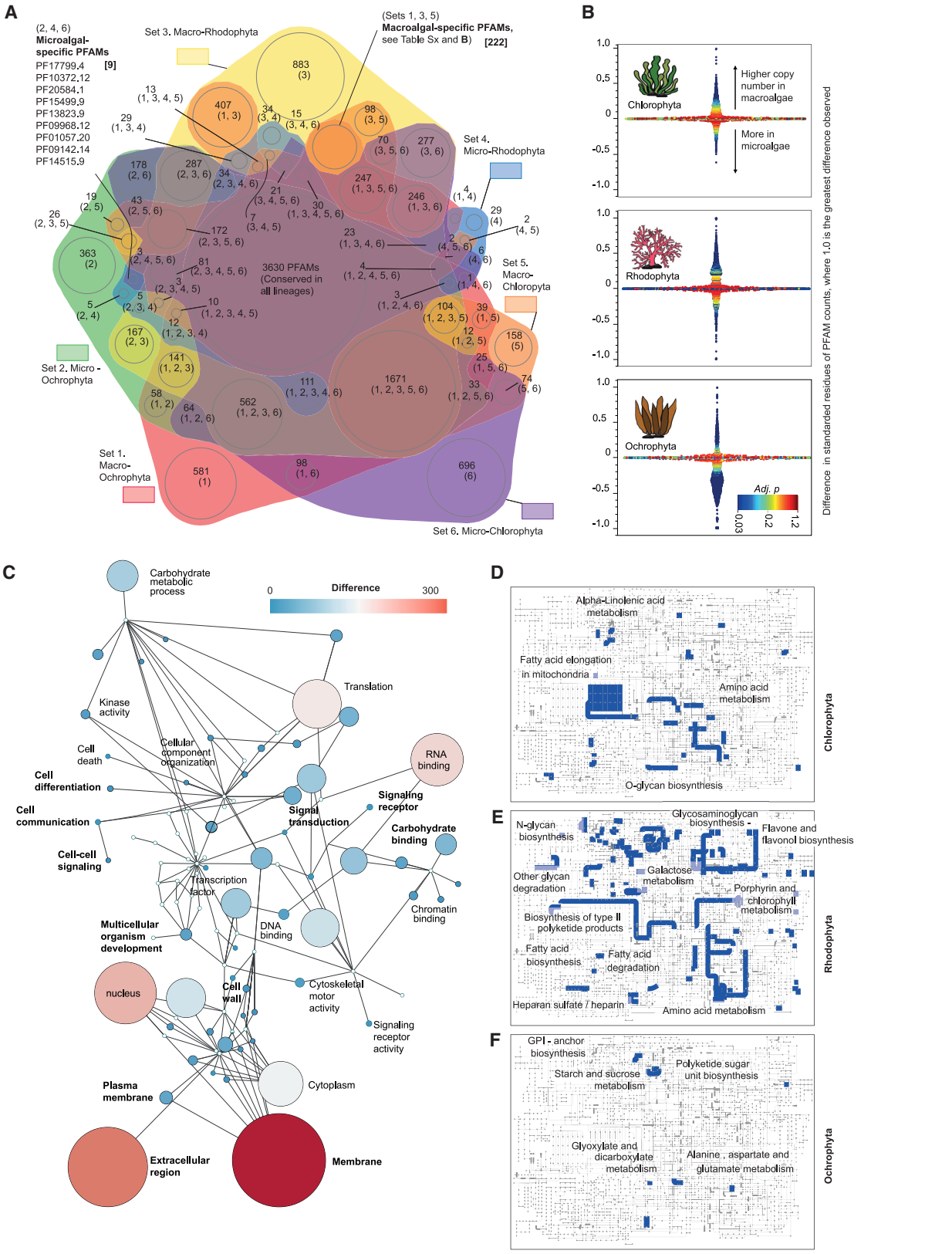
barycentric; three variables, represented by the average copy number of the PFAM, sum to a constant; bubble positions indicate copy number ratios (i.e., purity). Bubble size represents PFAM count totals for all phyla, where two is the smallest size, and 1000 is the largest.

(B) Selected enriched GO terms from PFAMs filtered at 80% purity ($n = 646$) in the Phaeophyceae ($n = 45$) with roles in maintaining a multicellular phenotype. PFAMs for cell organization and migration had high representation, alluding to roles for determining morphology and tissue organization in these phaeophyte algal genes. Contrasting with Ochrophyta and Chlorophyta, Rhodophyta ($n = 74$) displayed no GO enrichment in PFAMs with over 80% purity, possibly due to their longer evolutionary timeline and diversified traits, resulting in a generalist PFAM profile.

(C) PFAM domains of unknown function (DUFs) from each phylum as compared with their unicellular correlates using false discovery rate (FDR)-corrected batch *t-tests* in phyla-wide response screens. The x axis categories indicate whether a DUF was present at significantly different or equivalent levels or if statistical resolution was absent (inconclusive).

(D and E) Differential retention of key PFAMs in the three lineages. Each PFAM description includes multiple distinct PFAM domains differentially retained in the three lineages (see Supplemental Table 1). Most of the PFAMs distinguishing macroalgae from microalgae were predicted to function in the **(D)** membrane or **(E)** nucleus, outlining structural, signaling, and fundamental cell programming strategies toward maintaining multicellularity.

Please cite this article in press as: Nelson et al., Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity, Molecular Plant (2024), https://doi.org/10.1016/j.molp.2024.03.011



(legend on next page)

Please cite this article in press as: Nelson et al., Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity, Molecular Plant (2024), https://doi.org/10.1016/j.molp.2024.03.011

Hypergeometric tests were used to determine functional enrichment from unique and shared PFAM sets distinguishing macroalgae from their microalgal counterparts (Figure 3). The 222 macroalgal-specific PFAMs conserved in all three lineages were significantly enriched for 59 GO terms. Enriched GO terms related to morphology, extracellular matrix (ECM) formation, cell-cell adhesion, the establishment of cell polarity, and signaling were recovered (Figure 3C). For example, "binding" (GO:0005488, Z = 3.2, adj. p = 1.62 × 10^{-2}), "cell periphery" (GO:0071944, Z = 3.49, adj. p = 1.18 × 10^{-2}), "apical part of cell" (GO:0045177, Z = 3.88, adj. p = 2.15 × 10^{-2}), "cell differentiation" (GO:0030154, Z = 4, adj. p = 2.54 × 10^{-2}), "cytoskeleton organization" (GO:0007010, Z = 4.15, adj. p = 2.68 × 10^{-2}), "cell-cell adhesion" (GO:0098609, Z = 5.27, adj. p = 3.12 × 10^{-2}), "cell-cell recognition" (GO:0009988, Z = 9.11, adj. p = 2.54 × 10^{-2}), and "multicellular organismal process" (GO:0032501, Z = 3.64, adj. p = 2.54 × 10^{-2}) were GO terms associated with the enriched macroalgal PFAMs (See Supplemental Note 1A for expanded technical and contextual descriptions of PFAMs differentiating micro-from macroalgae and their GO enrichment). The high degree of enrichment in these proteins in macroalgae from any clade compared to their microalgal counterparts suggests they are associated with macroalgal multicellularity evolution. The macroalgal genes underlying the enrichment are responsible for many processes and functions known to be important for multicellularity in other lineages (Steinberg, 1975; Hedges et al., 2004; Ratcliff et al., 2012; Nagy et al., 2018; Strother et al., 2021).

The RRs of PFAMs in macroalgae (n = 124) were evaluated in a phyla-wide manner with batch *t-test*s (i.e., response screens) by comparison with microalgae from the corresponding clades (n = 125). We found 1059 instances of significantly different RRs (adj. p <0.05): chlorophytes had 10 losses and 61 gains, ochrophytes had 507 losses and 117 gains, and rhodophytes had 53 losses and 311 gains. There was a strong positive correlation between RR and clade age. Compared to microalgae, Chlorophyta and Rhodophyta macroalgae have expansions in a wide variety of gene families, indicating functional divergence after the development of multicellularity (Figure 3B). In contrast, unicellular stramenopiles had more diverse PFAMs than their macroalgal correlates (Figure 3B). The uniformity of coding potential in Ochrophyta macroalgae is likely due to their recent evolution from a common stramenopile ancestor (Cock et al., 2010a, 2010b).

Rhodophyta showed a drastic increase in the diversity and overall quantity of PFAM DUFs, while Chlorophyta and Ochrophyta had fewer DUFs compared to their unicellular correlates (Figure 2C, Supplemental Table 3). The predominant features of the annotated PFAMs, expanded in macroalgae, were membrane proteins (e.g., transporters and adhesome components). We used a set of 183 transporter PFAMs to detect any significant (adj p <0.05) gains or losses. Ochrophyte, but not chlorophyte or rhodophyte, macroalgae had significant expansions of a clan CL0184 transporter (PF07857, p = 1.32 × 10^{-13}, Supplemental Table 3). Of these, 29 had significantly different retention in macroalgae compared to microalgae. An endoplasmic reticulum vesicle transporter (PF07970) was significantly expanded in Rhodophyta but not Chlorophyta or Ochrophyta macroalgae (p = 2.20 × 10^{-234}). A potassium (K^+) transporter (PF02705), widely conserved in microalgae from all three phyla, was missing from chlorophyte and rhodophyte but retained in ochrophyte (eq. p = 1.95 × 10^{-270}) macroalgae. This K^+ transporter may have been a new introduction to microalgae after the major rhodophyte and chlorophyte seaweeds emerged, or it could have been lost from the older two macroalgal lineages.

Several other transporters were expanded in one or more macroalgal lineages. A branched-chain amino acid transporter was significantly expanded in all three macroalgal phyla compared to their unicellular correlates (PF02653; Chlorophyta p = 8.99E−227, Ochrophyta p = 1.62E−30, Rhodophyta p = 2.48E−88). The need to transport amino acids is inherent in multicellular systems where nutrient absorption and energy expenditure are spatially segregated. A chromate transporter (PF02417) was expanded in Rhodophyta (p = 1.25E−19), but not in Chlorophyta or Ochrophyta macroalgae compared to their unicellular correlates. A periplasmic zinc uptake complex component (ZnuA, PF01297, chelatase)

Figure 3. Shared and unique PFAMs in micro- and macroalgae.

(A) Intersection plot showing shared and unique PFAMs in chlorophyte, rhodophyte, and ochrophyte micro- and macroalgae. Only nine PFAMs were exclusively conserved in chlorophyte, rhodophyte, and ochrophyte microalgae (listed); 222 PFAMs were exclusively conserved in the three macroalgal lineages. Rhodophyta macroalgae had the most unusual PFAMs (883) compared to the other groups.

(B) Protein family counts per macroalgal genome were compared for each PFAM domain (n = 12 938 PFAMs in 38 814 comparisons, Supplemental Table 3) between macroalgae and their microalgal counterparts in each phylum using FDR-corrected batch *t-test*s in phyla-wide response screens. Unicellular ochrophytes had a greater diversity of PFAMs than their multicellular counterparts, correlating with the short period in which the latter had to evolve.

(C) Unique PFAM sets were used in domain-centric enrichment analyses. No GO terms were significantly enriched in most PFAM sets defined by phyla and multicellular phenotypes. The 222 macroalgal-specific PFAMs were used in dcGO (Fang and Gough, 2013) enrichment analyses to determine functional enrichment (GO terms; see also Supplemental Table 2). Enrichment of cell morphology, specialization, and multicellular organism organization GO terms was found in this set. The PFAMs constituting each GO term essentially comprise core gene collections unique to multicellular aquatic algae compared to their unicellular relatives but conserved in all three major phyla. Enrichment Z scores were plotted to Cytoscape network nodes in a model reconstructed from the relationships in the go.obo GO infrastructure file (https://doi.org/10.5281/zenodo.2529950). The Cytoscape network used to create this panel is in Supplemental Data 3 (.cys file).

(D) KEGG map of chlorophyte macroalgal-specific metabolic pathways. The PFAMs varying with robust significance between macroalgae of the three clades compared to their microalgal counterparts were used to find Enzyme Commission (EC) (McDonald and Tipton, 2023) codes and mapped to pathways in the Kyoto Encyclopedia for Genes and Genomes (KEGG) using iPATH3 (Letunic et al., 2008). The EC lists in Supplemental Table 3 can be processed in iPath3 (https://pathways.embl.de) (Letunic et al., 2008) for an interactive exploration of macroalgal-specific reactions.

(**E**) KEGG map of rhodophyte macroalgal-specific metabolic pathways.

(**F**) KEGG map of ochrophyte macroalgal-specific metabolic pathways.

Please cite this article in press as: Nelson et al., Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity, Molecular Plant (2024), https://doi.org/10.1016/j.molp.2024.03.011

was similarly expanded in Rhodophyta but not in Chlorophyta or Ochrophyta macroalgae compared to their unicellular correlates (adj. $p = 9.23 \times 10^{-11}$). This observed expansion of a zinc chelation gene in Rhodophyta may be responsible for the increased zinc concentration found in red, but not green or brown, seaweeds (Stengel et al., 2004). Macroalgae can accumulate heavy metals from their environment. This accumulation is a result of various mechanisms, including adsorption on the cell surface and intracellular uptake, and has supported seaweed-based bioremediation efforts. Red macroalgae have already achieved moderate success in removing unwanted inorganic molecular (e.g., heavy metals) from sensitive ecosystems and industrial processes (He et al., 2008; Sode et al., 2013; Kidgell et al., 2014; Roberts et al., 2015; Deniz and Ersanli, 2018; Elizondo-Gonzalez et al., 2018; Sun et al., 2019; Luo et al., 2020; Astrahan et al., 2021; Mansour et al., 2022; Znad et al., 2022). Our results show two expanded gene families in Rhodophyta macroalgae that could boost their bioremediation potential.

Metabolic pathways gained in macroalgae during their evolution of multicellularity

Macroalgae have evolved novel developmental programs coordinating morphogenesis. Cell polarity is established, and cell differentiation produces cell types underlying novel multicellular traits. For example, the millimeter-long holdfast rhizoid cells in Rhodophyta species (e.g., *Porphyra* [Brawley et al., 2017] and *Griffithsia* [Goff and Coleman, 1987]) facilitate survival in violent intertidal areas. Other cell-type specializations, such as the mat-forming stolons of *Caulerpa cylindracea*, promote invasiveness by smothering and over-crowding other native macroalgae (Katz et al., 2021). Here, we show unique metabolic pathways distinguishing macroalgae from their unicellular relatives (Figure 3D–3F, Supplemental Table 3). A striking feature in metabolic maps of macroalgal-specific genes is the prevalence of polysaccharide metabolism pathways. These pathways contribute to complex, tissue-specific glycoprotein patterns that define organismal morphology and cell-development processes. Macroalgae rely on diverse sugars and sugar derivatives for specific ECM components (Domozych and Domozych, 2014; Kloareg et al., 2021; Mazeas et al., 2022). Our data revealed lineage- and habitat-specific sugars and broadly conserved sugar metabolism genes. Rhodophyte macroalgae exhibited significantly more glycosyl hydrolases and carboxylesterases than their microalgal correlates (adj. $p < 0.05$), while carbonic anhydrases were more abundant in macroalgal chlorophytes (PF00194, adj. $p < 0.05$), likely for concentrating CO_2 in Ulvophyceae. See Supplemental Note 1B for expanded technical and contextual descriptions of macroalgal ECM analyses.

ECM polysaccharides play crucial roles in cell adhesion mechanics and provide protection for macroalgae from their surrounding environment. Components of seaweed ECMs, such as mannose derivatives, sulfated polysaccharides (SPs), and acidic-residue-rich linking proteins such as integrins, are indispensable for macroalgae, and sets of these genes distinguish them from microalgae and land plants. SPs, which are particularly structurally diverse and unique to some marine plants and macroalgae, comprise numerous building blocks for tissue-specific ECMs. These adhesive molecules have been implicated for use in human bone tissue regeneration (Venkatesan et al., 2019). Genes involved in mannose, another important ECM component, metabolism were conserved across all three macroalgal phyla examined, suggesting their significance for aquatic multicellular algal survival. Macroalgae lack glycosyltransferase activity required in complex-type N-glycan biosynthesis (Yoshiie et al., 2012). Mannose, and other osmolyte-type sugars, can also promote marine survival through conferring resistance to salinity. Although widely conserved in seaweeds, sparse examples of the mannose metabolism's role in salt stress resistance in land plants have been reported (Hirano et al., 2000; Ma et al., 2014; Yu et al., 2021). Thus, the mannose metabolic pathways reported here may be of value for crop-improvement strategies.

Macroalgal-specific PFAMs shared in all three lineages reveal convergence toward these functions (Figure 3). We used this set of 222 PFAMs as queries in a dcGO enrichment analysis and found enrichment in GO terms related to cell membrane reorganization, cell polarity establishment, cytoskeleton rearrangement, and cell specialization in vascular tissue (Supplemental Table 2). Genes related to mannose and sulfated polysaccharide metabolism also distinguished them from microalgae. We found a significantly higher number of sulfatase domains in rhodophyte macroalgae than in microalgae (PF00884, clan = CL0088, adj. $p = 1.43 \times 10^{-6}$, Supplemental Table 3). Land plants generally lack SPs and high-mannose cell walls. However, angiosperm seagrasses such as Hydrocharitaceae, Ruppiaceae, Posidoniaceae, Cymodoceaceae, and Zosteraceae contain both SPs and high-mannose-type N-linked glycans. The convergent evolution of these ECM components among photosynthetic, marine, and multicellular eukaryotes from diverse lineages implies their adaptiveness in a marine environment. Conversely, Rhodophyta display remarkable complexity in ECM-related pathways, setting them apart from the other two lineages. These ECM gene collections from Rhodophyta macroalgae were some of the most prominent examples of evolutionary divergence seen in our datasets. Presumably, the much longer evolutionary period of the Rhodophyta macroalgal clades allowed for the development of more complex ECM biosynthesis gene repertoires. We speculate that the greater number of glaciation cycles experienced by the Rhodophyta drove the higher-complexity ECM systems inferred from our data, as an adaptation to freeze-thaw cycles requires flexible cell-cell adhesive networks.

Macroalgal multicellularity facilitated by cell polarity, communication, and adhesion

Macroalgae exhibit complex life cycles that vary among species. Pioneering research into the life cycles of model phaeophyte macroalgae, such as *Ectocarpus* sp. (Muller, 1966, 1967b, 1967a, 1968) and members of the Fucales (Vreeland et al., 1993; Kropf, 1997) and Dictyotales orders (Charrier et al., 2017), has uncovered intricate developmental strategies. Typically, macroalgae progress through four multicellular stages, including both fertile and vegetative gametophytes and sporophytes (Yang et al., 2016; Macaisne et al., 2017; Arun et al., 2019). In some taxa, sexual reproduction polarity is induced at the zygotic phase (Kropf, 1997), and polarized embryos develop following zygote fertilization (Charrier et al., 2017). One key feature of macroalgal cells distinguishing them

Please cite this article in press as: Nelson et al., Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity, Molecular Plant (2024), https://doi.org/10.1016/j.molp.2024.03.011

from microalgae is their extensive cell-cell communication networks. Our dataset suggests that these networks partially arose due to gene gains compared to their unicellular ancestors. The thrice-conserved, macroalgal-specific 222 PFAM set showed dcGO enrichment for "cell-cell recognition" (GO:0009988, $Z = 9.11$, adj. $p = 2.54 \times 10^{-2}$; Supplemental Table 2), which included an FAD-dependent oxidoreductase (IPR006076). Various redox processes are involved in cell-cell communication and are thought to have been essential in animal multicellularity development (Blackstone and Ellison, 2000). Thiol-based sensors constitute central machinery in plant defense signaling networks (Chae et al., 2023). We observed significant expansions in copper zinc superoxide dismutase (PF00080, Sod_Cu) in Rhodophyta (adj. $p = 4.81 \times 10^{-6}$) but not in Chlorophyta or Ochrophyta macroalgae (see also Supplemental Table 3). Ochrophyta, but not Rhodophyta or Chlorophyta, showed significant expansion of a thioredoxin gene (PF00085, $p = 1.9 \times 10^{-5}$). These distinct expansions in the Rhodophyta and Ochrophyta macroalgae highlight non-plant mechanisms for redox sensing in multicellular phototrophs.

Macroalgae exhibit complex cell-to-cell adhesion, facilitated by elaborate ECMs or direct connections, a hallmark of multicellular development. Proteins and polysaccharides, collectively known as the adhesome, mediate these processes (Whittaker et al., 2006; Zaidel-Bar et al., 2007,; Zaidel-Bar and Geiger, 2010; Zaidel-Bar, 2013; Winograd-Katz et al., 2014). Coordinated cell-to-cell adhesion, facilitated by complex ECMs or direct connections, is another hallmark of multicellular development (Domozych and Domozych, 2014; Brunet and King, 2017; Kloareg et al., 2021). Variations in cell-cell adhesion mechanisms have been proposed to underscore the progression from simple to complex multicellularity (Strother et al., 2021), and our data support this hypothesis. Thus, we interrogated our genome datasets for genes involved in adhesion that could facilitate multicellularity, especially those implicated in developmental and cellular differentiation processes. The multistage adhesive process starts during the early development of zygotes and influences the survival rates and reproductive success of macroalgae (Vreeland et al., 1993; Matt and Umen, 2016), We defined the macroalgal adhesome as including cadherin, integrin, armadillo (ARM), lectin, F actin-binding calponin homology, Kringle, Pleckstrin homology (PH), Src homology 2 (SH2), and LIM PFAMs (n = 110 PFAMs; see Supplemental Table 4). Other enzymes, such as guanine nucleotide exchange factors and GTPase-activating proteins, protein tyrosine, serine/threonine kinases and phosphatases, and E3 ligases and proteases, regulate adhesion by post-translational modification and do not directly adhere to other proteins. Differences in levels of these proteins in macroalgae compared to microalgae were seen in our response screen (Supplemental Table 3); here, we focused on PFAMs with the potential to bind to proteins presented by other cells or in the ECM (i.e., cell–cell and cell–matrix) and their accessory domains, or codomains, which regulate and fine-tune adhesive interactions (Zaidel-Bar and Geiger, 2010; Zaidel-Bar, 2013; Collins et al., 2017; Hernandez-Vasquez et al., 2017; Thompson et al., 2021). From 110 PFAMs constituting the queried adhesome (Supplemental Table 4), 146 539 PFAMs in 124 queried macroalgal genomes, or ~1154 adhesome PFAMs/genome, were recovered (Supplemental Table 4). Several macroalgal cell-cell linkage and communication genes were either absent or present in higher copy numbers than microalgae (Supplemental Table 4). Our adhesome analyses point to a significant role for these genes/domains in the evolution of multicellularity in macroalgae.

We analyzed genes containing these domains for codomains. Altogether, 7065 adhesome codomains or ~55.6 adhesome codomains were detected per genome. These codomains likely extend the functionality of their proteins, providing combinatorial complexity to overcome a wide variety of biotic and abiotic challenges faced by their host lineages. We compared the adhesome across macroalgal phyla, native climates, and habitats (Figure 4). We saw 294 adhesome proteins that significantly differed in copy number between phyla and 770 that significantly varied by climate (adj. $p < 0.05$). The 5365 significant differences in adhesome PFAMs by phyla within habitats (Supplemental Table 4) suggest that environmental cues had shaped adhesome development in a convergent manner in the three macroalgal phyla. Still, we found that 129 PFAMs significantly differed among phyla within their respective habitats. These lineage-specific adhesome components detail the genetic blueprints for multicellular development in the rhodophyte, chlorophyte, and ochrophyte seaweeds. Rhodophyta species had more extensive collections of adhesome elements, and subpolar species from any clade had substantially less. A minimal adhesome gene set in subpolar algae suggests that fewer genes for these proteins may be necessary in subpolar regions.

The compendium of cadherin codomains seen here suggests macroalgae exploit the capacity of this network to use mechanisms outlined in the differential adhesion hypothesis (Wiseman, 1977; Foty and Steinberg, 2005; Amack and Manning, 2012) for complex tissue and body plan organization as well as climate-specific adaptation (see also Supplemental Table 4). For example, epiphytic phaeophyte macroalgae had significantly more Laminin_G_3 domains than epiphytic rhodophytes (adj. $p = 0.03$). Phaeophyte algae had 58 adhesome domains with significantly different levels than the other phyla, with 69 comparisons showing significantly fewer adhesome domains than other clades. Ochrophyta macroalgae had adhesome gene expansions, including only five PFAMs. These consisted of the intraflagellar transport 81 calponin homology domain (IFT81_CH, PF18383, increased in phaeophyte compared to [1] chlorophyte epiphytes, adj. $p = 1.07789 \times 10^{-15}$; and [2] rhodophyte epiphytes, adj. $p = 3.26949 \times 10^{-26}$), calponin homology (CH) domain (PF00307, increased in mangrove-dwelling phaeophytes compared to chlorophytes, adj. $p = 0.0008$), and the concanavalin A-like lectin/glucanases (Laminin_G_3, higher in Rhodophyta [epiphytic], adj. $p = 0.031$).

Comparisons of adhesome domains in the three macroalgal clades compared to their microalgal counterparts showed 191 significant differences (adj. $p < 0.05$). These results outline various mechanisms for tissue formation and organization of multicellular, marine-dwelling photosynthetic eukaryotes. Chlorophytes (n = 11) had 15 adhesome domains with significantly different levels than their microalgal correlates. Of these, 14 were present at increased levels. An acetamidase formamidase (FmdA_AmdA, PF03069) adhesome codomain was present at significantly fewer levels in macroalgae, and this may represent a reduced need

Please cite this article in press as: Nelson et al., Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity, Molecular Plant (2024), https://doi.org/10.1016/j.molp.2024.03.011

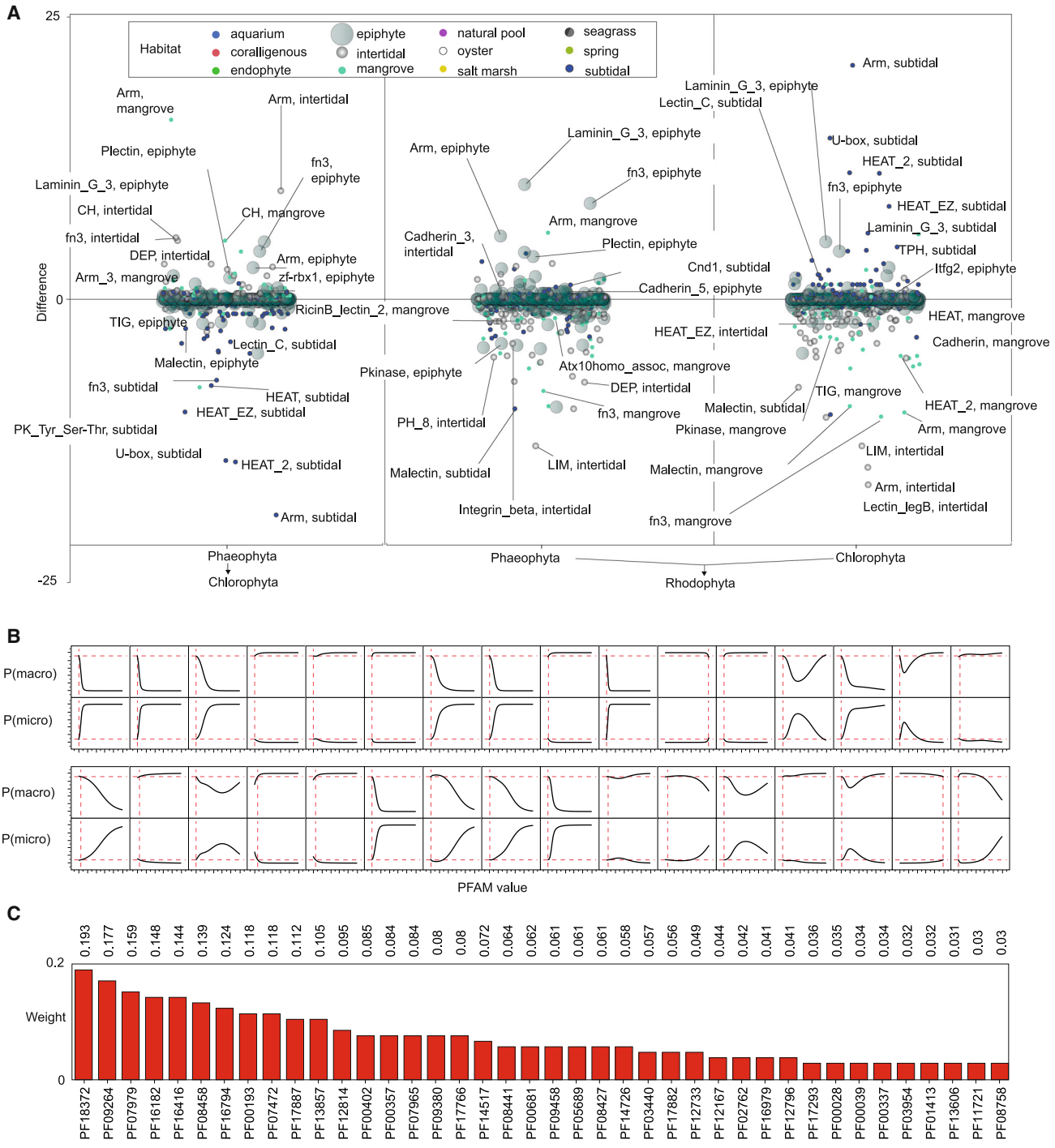


Figure 4. The macroalgal adhesome as a function of phyla, habitat, and multicellularity.

(A) Macroalgal adhesome PFAMs were compared across phyla and habitats (n = 124 genomes) in FDR-corrected batch *t-tests* (Supplemental Table 4). The PFAMs selected from Supplemental Table 1 comprising the adhesome were cadherin, integrin, armadillo (ARM), lectin, F actin-binding CH, Kringle, PH, Src homology 2 (SH2), and LIM PFAM domains, altogether consisting of 110 distinct HMM-based PFAM domains. All data points shown are significantly different at adj. p <0.05. Groups of macroalgae listed in the top boxes were compared to those listed along the x axis, where a positive number signifies an increase in the top level.

(B) Prediction profiles for whether a subject is microalgal or macroalgal as a function of adhesome PFAMs. These variables contribute the most toward classification in a deep neural network fitting adhesome PFAMs from the macro- and microalgal genomes. The charts are marginal plots describing the effect of increasing or decreasing PFAM domain counts in the macroalgal groups as predictive of either macro- or micro-cellularity status.

(C) Attribute influence (i.e. variable importance) shows the predominant decision-making adhesome PFAMs defining macroalgal multicellularity. These adhesome constituents served as adequate markers for the transition to multicellularity in macroalgae.

Please cite this article in press as: Nelson et al., Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity, Molecular Plant (2024), https://doi.org/10.1016/j.molp.2024.03.011

to assimilate nitrogen from these sources in multicellular chlorophytes. Ochrophyta macroalgae (n = 45) had the most significantly differing adhesome codomains (116 comparisons with adj. p <0.05) compared to their unicellular ochrophyte correlates. Differing domains included two immunoglobulins (Ig_3, PF13927 and Ig_2, PF13895) as well as various TFs and kinase adhesome codomains. For example, Ochrophyta macroalgae had significantly more diaphanous FH3 domains (Drf_FH3, TPR-like, clan = CL0020, PF06367). This domain is a core component of actin regulatory genes and has a wide-reaching influence over cell morphology and adhesive properties.

Macroalgal adhesome domain composition (including cadherin, integrin, ARM, lectin, F actin-binding CH, Kringle, PH, SH2, and LIM PFAMs; n = 110 PFAMs) was sufficient to build an artificial neural network (aNN) that could accurately determine the unicellular or multicellular status of an alga (R^2 = 92.9; Figure 4, Supplemental Figure 4). This aNN was composed of 24 layers of 13 nodes, each consisting of eight tanH, four linear, and one Gaussian activation function (8-4-1 TanH-L-Gx24). The predictive capacity of these PFAMs in our aNN point to macroalgal adhesomes being critical factors for their multicellularity status.

Cadherins had very heterogeneous distributions in the macroalgal genomes, with only 85 out of 124 genomes surveyed containing at least one cadherin domain and some species showing substantial expansions in cadherins. The sparsity shows that cadherins may not be necessary for a multicellular phenotype in macroalgae. This finding contrasts with studies showing that cadherins alone are capable of driving complexity in emerging, evolving multicellular lineages (Foty and Steinberg, 2005) and indicates that a wide variety of cell-cell adhesion mechanisms can support a multicellular lifestyle. Adhesome networks of cadherins, collectively known as the cadhesome (Zaidel-Bar, 2013), are generally characterized by many more linkers than nodes. Non-cadherin PFAMs accompanying cadherin domains in the macroalgal genes coded for diverse functions. Many cadherin codomains had other similar binding and receptor functions. Cohesins, Kringle domains, Dockerins, Sushi domains, the bilobal receptor L domain, and various lectins were cadherin codomains. These codomains play dominant roles in cell-cell adhesion dynamics and intracellular communication, including mechanotransduction (Campbell and Humphries, 2011; Tong et al., 2018; Zheng and Leftheris, 2020) and small-molecule sensing (Forzisi and Sesti, 2022). Some macroalgal cadherins had TIG/IPT domains characterized by immunoglobulin-like folds. These TIG domains are commonly found in cell surface receptors such as Met and Ron and DNA-binding TFs (Aravind and Koonin, 1999; Bork et al., 1999).

Differential binding from cadherins, integrins, and other adhesive proteins in differentiated cell types is the basis for the differential adhesion hypothesis (Steinberg, 1975; Wiseman, 1977; Nicol and Garrod, 1979; Thomas and Yancey, 1988; Foty and Steinberg, 2005; Emily and Francois, 2007; Amack and Manning, 2012). This hypothesis outlines mechanisms for differentiated cell sorting, tissue morphogenesis, rearrangements, spreading, and segregation. It posits that cadherins' type, activation state, and density directly mediate cellular organization and organism body plan by minimizing adhesive-free energy through optimizing cell-to-cell binding (Foty and Steinberg, 2005). Thus, different types of cadherins in the macroalgal genomes likely play fundamental roles in their respective lineages' tissue morphogenesis and overall morphologies. We analyzed the macroalgal cadherins and other adhesome components for other concurrent PFAM domains to identify lineage- and species-specific cell connection-related PFAMs that may be responsible for their multicellular morphogenic body plans. See Supplemental Note 1C for expanded descriptions of cadherin and other adhesome codomains in the macroalgal genomes.

Changes in integrin and cadherin rigidity sensing regulate cell adhesion properties (Collins et al., 2017). This rigidity modulation regulates actin organization and membrane dynamics during cell-cell adhesion. Different domains in these proteins could alter how mechanotransduction signals are interpreted (e.g., through TF signaling) and influence cellular polarity determination. For example, vast hierarchies of TFs regulate macrophage polarization in metazoans (Czimmerer and Nagy, 2023). In instances where cadherin codomains bind specific macromolecules, such as Kringles, which bind to cell membranes, the cadherins and integrins can more precisely determine the organisms' physiology by requiring a body plan that conforms to its macromolecular binding requirements. Dynamic cellular adhesion by the "mix-and-match" design might facilitate resilient survival in seaweeds clinging to intertidal rocks despite violent waves and daily dehydration in baking sunlight. The evolution of integrins and cadherins with exotic codomains could provide unexpected functional benefits through far-reaching consequences in an organism's body plan and resultant morphology. Our data reveal the combinatorial adhesome complexity facilitating the evolution of novel multicellular phenotypes in three macroalgal lineages.

Macroalgal endogenous viral elements and concurrent functional domains

The interaction of early eukaryotes with viruses has been hypothesized to favor the evolution of multicellularity as a defense against infection (Iranzo et al., 2014). Here, we examined the role of viruses in the evolution of macroalgal multicellularity, not through selection for immune systems but as a means of exchanging genetic information among lineages and conferring gain-of-function mutations. We examined viral elements in the coding sequences (CDSs) from the macroalgal genomes using HMMsearch (Potter et al., 2018) as in Skewes-Cox et al. (2014) and Nelson et al. (2021). The absolute quantities of their inclusion were not necessarily clade dependent, as clades had highly heterogenous inclusion of viral families (VFAMs), and no phyla surveyed had significantly different quantities of VFAMs from the other two (Figure 5, Supplemental Table 5, VFAMs). We observed clade-dependent (i.e., high conservation, retained by most members) and clade-independent (e.g., sporadic or correlating with native climate) virus sequence integration in the macroalgal genomes (see Supplemental Note 1D for expanded technical and contextual description of the global VFAM analyses).

EsV-1-7-like domains are sporadically found throughout eukaryotes; previously, only two chlorophyte species and one cryptophyte species were known to share EsV-1-7-like domains (Macaisne et al., 2017). EsV-1-7 proteins represented

Please cite this article in press as: Nelson et al., Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity, Molecular Plant (2024), https://doi.org/10.1016/j.molp.2024.03.011

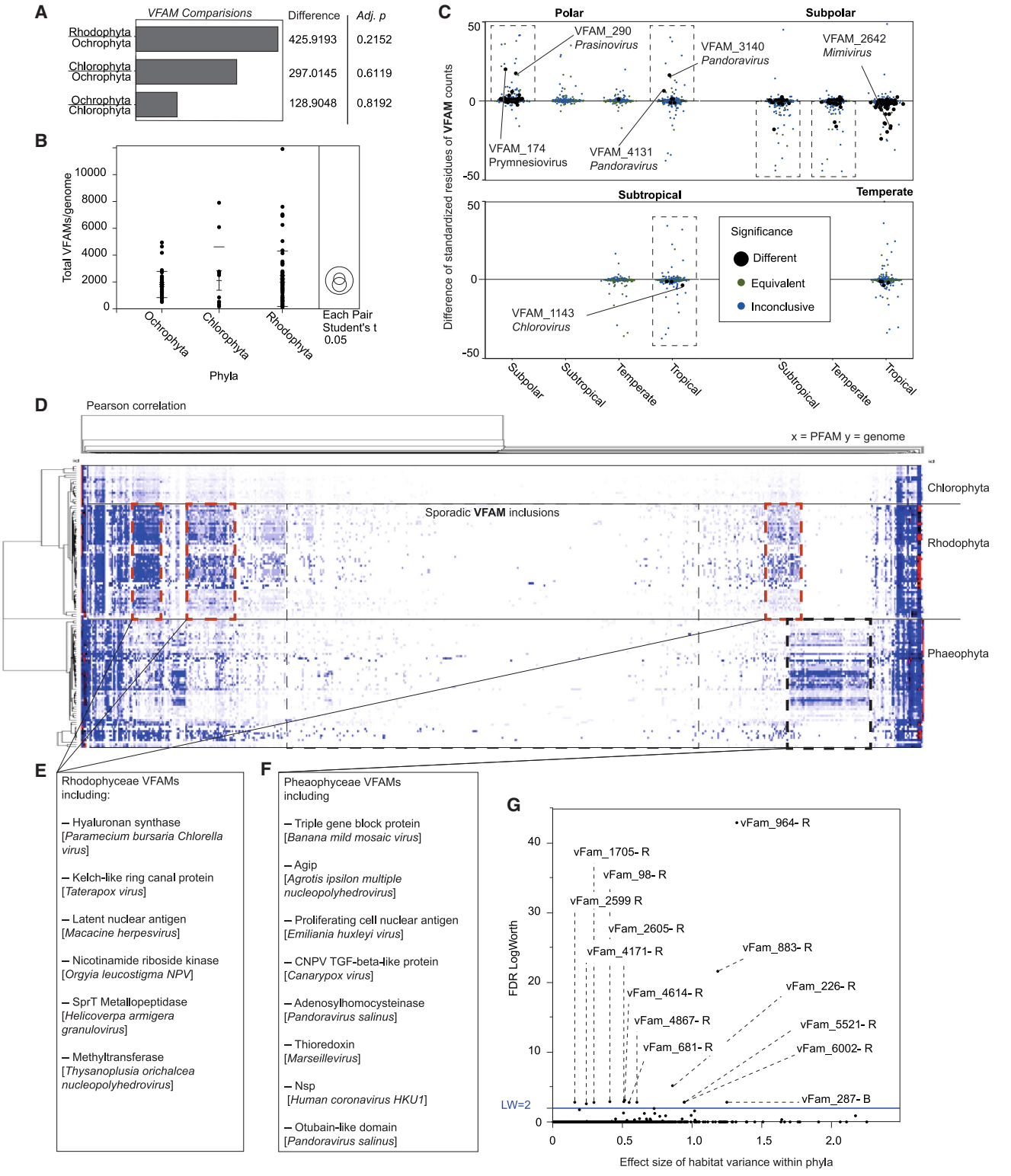


Figure 5. VFAM sequences in macroalgal genomes and their variance with clade and climate.

(A and B) Comparison of total VFAMs in macroalgal genomes from Rhodophyta (*n* = 74), Ochrophyta (*n* = 43), and Chlorophyta (*n* = 11). Rhodophyta had the highest average VFAMs per genome, followed by Chlorophyta, then Phaeophyceae. No phyla significantly differed from another in terms of average VFAMs per genome.

(C) Comparison of average VFAMs per genome by native climates. Data points represent genome differences in standardized residues of VFAM means compared between the climates listed. Large black dots represent significant differences (adj. p <0.05).

(legend continued on next page)

Please cite this article in press as: Nelson et al., Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity, Molecular Plant (2024), https://doi.org/10.1016/j.molp.2024.03.011

the highest copy number PFAM for the Ochrophyta macroalgal genomes differentiating them from the other phyla (Figure 2A, largest bubble in Ochrophyta >80% purity). Phaeophytes had 293.5 ± 229.4 EsV-1-7 domains per genome (n = 43). The high Esv-1-7 counts in Phaeophyceae compared to the very low or nonexistent EsV-1-7 domains in other macroalgae establish Esv-1-7 motif retention as a defining feature for the clade. Four Rhodophyta species were also found with about five EsV-1-7 motifs per genome: *Porphyra plocamiestris* (CCMP673), *Catenella* sp. (CCMP2858), *Porphyra miniata* (CCAP1379/2), *Caloglossa vieillardii* (CCMP3076), and *Polysiphonia violacea* (CCAP1348/5). The connection of Esv-1-7 domains and various proteins necessary for a multicellular phenotype suggests that the acquisition of lysogenic virus-sourced genes comprises essential genomic architecture in the Phaeophyceae.

We surveyed genes with EsV-1-7 domains for accessory domains, or codomains (Figure 6). Altogether, 931 codomains were found in 13 696 EsV-1-7 motif-containing genes (n = 43 phaeophyte algal genomes surveyed). Codomains included the Rieske domain (PF00355.2, E = 7.9 × 10^{-8}), which is involved in electron transfer in many systems (Dong et al., 2019; Feyza Ozgen et al., 2020; Mei et al., 2020; Nji Wandi et al., 2020; Purhonen et al., 2020; Hou et al., 2021; Di Trani et al., 2022). The acquisition of Rieske domains and their long-term maintenance indicates a fundamental shift in the biological capacity of the host lineage. Retained EsV-1-7 intact codomains may offer new metabolic possibilities for their hosts. For example, Rieske proteins have conferred photosynthesis to non-photosynthetic cells (Feyza Ozgen et al., 2020). Such a drastic phenotype change was observed experimentally, but the long-term geological implications of massive arrays of viral-sourced genes in macroalgal lineages have not yet been appreciated. We found Rieske-Esv-1-7 codomains in 23 of the 43 Phaeophyceae genomes, suggesting that the function of this viral-sourced gene is partially conserved. See Supplemental Note 1E for an expanded description of EsV-1-7 codomains and gene structures.

The other prominent viral-origin domain found in Phaeophyceae genomes was the FNIP motif. We saw 4478 FNIP domains in phaeophyte algal genomes, with 195 accessory codomains sporadically distributed throughout genes. Plectins comprised 44 FNIP codomains, suggesting a conserved role for FNIP-Plectins. Plectins link actin microfilaments, microtubules, and intermediate filaments and maintain the mechanical integrity and viscoelasticity in tissues (Broussard and Green, 2022; Marks et al., 2022; Prechova et al., 2022; Wiche, 2022). Protein kinases were also FNIP codomains in phaeophyte genomes. A protein tyrosine and serine/threonine kinase (PF07714) and the protein kinase PF00069 were FNIP codomains. As a representative example, gene 4491 of scaffold 32 in the *Ectocarpus siliculosus* genome (Cock et al., 2010b) had FNIP motifs adjacent to a protein tyrosine and serine/threonine kinase (PF07714) domain. These types of genes were conserved endogenous viral elements in Phaeophyceae genomes. Protein tyrosine and serine/threonine kinases were some of the most significantly varying endogenous viral-origin proteins (EVOPs) among native macroalgal climates; they may provide climate-specific fitness.

The ratio of nonfunctional to functional elements accompanying EsV-1-7 domains was determined in selected genomes (see Supplemental Table 5). The phaeophyte alga *Botrytella uvaeformis* (CCAP1311/1) had 491 Esv-1-7 domains. Only six EsV-1-7 codomains were found: two Rieske (PF00355.29) domains, a zinc-finger (zf-MIZ, PF02891.23), an E3_ubiquitin ligase (RBR, PF18091.4), and two DUF5025 (PF16428.8) domains. This ratio implies a high number of viral insertions, but few of these sequences had known functions. Similar conclusions could be made from the manual inspection of ∼800 Interproscan search results for EsV-1-7 genes with codomains. Only ∼5% of these genes had defined functions. Most were characterized by tracks of EsV-1-7 motifs (5–20 copies of each protein) and a short disordered or predicted transmembrane region at either terminus. EsV-1-7 motifs were found on predicted intracellular and extracellular domains, suggesting that these proteins have signaling roles or may even function as structural proteins in Phaeophyceae.

Strikingly, we also observed EsV-1-7 cysteine-rich repeats as FNIP codomains. The FNIP motif is not coded for by EsV-1 and is instead found in members of the *Megaviradae*. Their coexistence in an open reading frame (ORF) alludes to a chimeric origin. The *Megavirus baoshan* has 82 protein-coding genes containing FNIP repeats (Xia et al., 2022). Most of these FNIP-containing proteins had terminal F-box domains, and up to 28 FNIP copies are in each gene. Interactome analysis of the FNIP protein Mb0983 showed how *M. baoshan* exploits its host's ubiquitin-proteasome pathways during infection. Specifically, FNIP-Mb0983 interacts with the GTPases Rap1B and Rab7A in the cytoplasm of *Acanthamoeba castellanii* (Xia et al., 2022). Thus, a viral-sourced, permanently acquired reprogramming of cell fate, regulated in a dose-dependent or on/off manner, could lead to differentiated cell types. Our data also suggest that megaviruses have provided phaeophyte algae with plectin genes.

Metabolic complexity appears to have increased in all the major macroalgal lineages, partly because of virus gene donation.

(D) Viral families (VFAMs, x axis) in species (y axis) from each of the three phyla. VFAMs were hierarchically bi-clustered using average Euclidean distance; the Z score correlates with relative VFAM counts. A cluster of VFAMs mostly unique to phaeophyte algae are shown in the dashed line box. These VFAMs are also found in viruses infecting Rhodophyta and Chlorophyta algae, which hints at cross-phyla infection; however, these specific VFAMs were not in the Rhodophyta or Chlorophyta algal genomes.

(E) Rhodophyta showed two major clusters of VFAMs, mostly unobserved in the other phyla.

(F) Phaeophyceae showed two major clusters of VFAMs, mostly unobserved in the other phyla.

(G) Effect size and FDR logworth of habitat variance within phyla on VFAM contents, where logworth = −log10(p value) and the effect size is the normalized mean of the sum of squares. The strongest effect sizes were seen in Rhodophyta VFAMs, which can be explained by the greater time over which they evolved niche-specific adaptations. VFAMs highlighted here include VFAM 1705 (Prasinovirus), VFAM_964 (Ictalurivirus), and VFAM_681 (Orbivirus). Extended descriptions of VFAMs from these analyses are in Supplemental Table 5, Supplemental Data 3, and are also at derisilab.ucsf.edu.

Please cite this article in press as: Nelson et al., Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity, Molecular Plant (2024), https://doi.org/10.1016/j.molp.2024.03.011

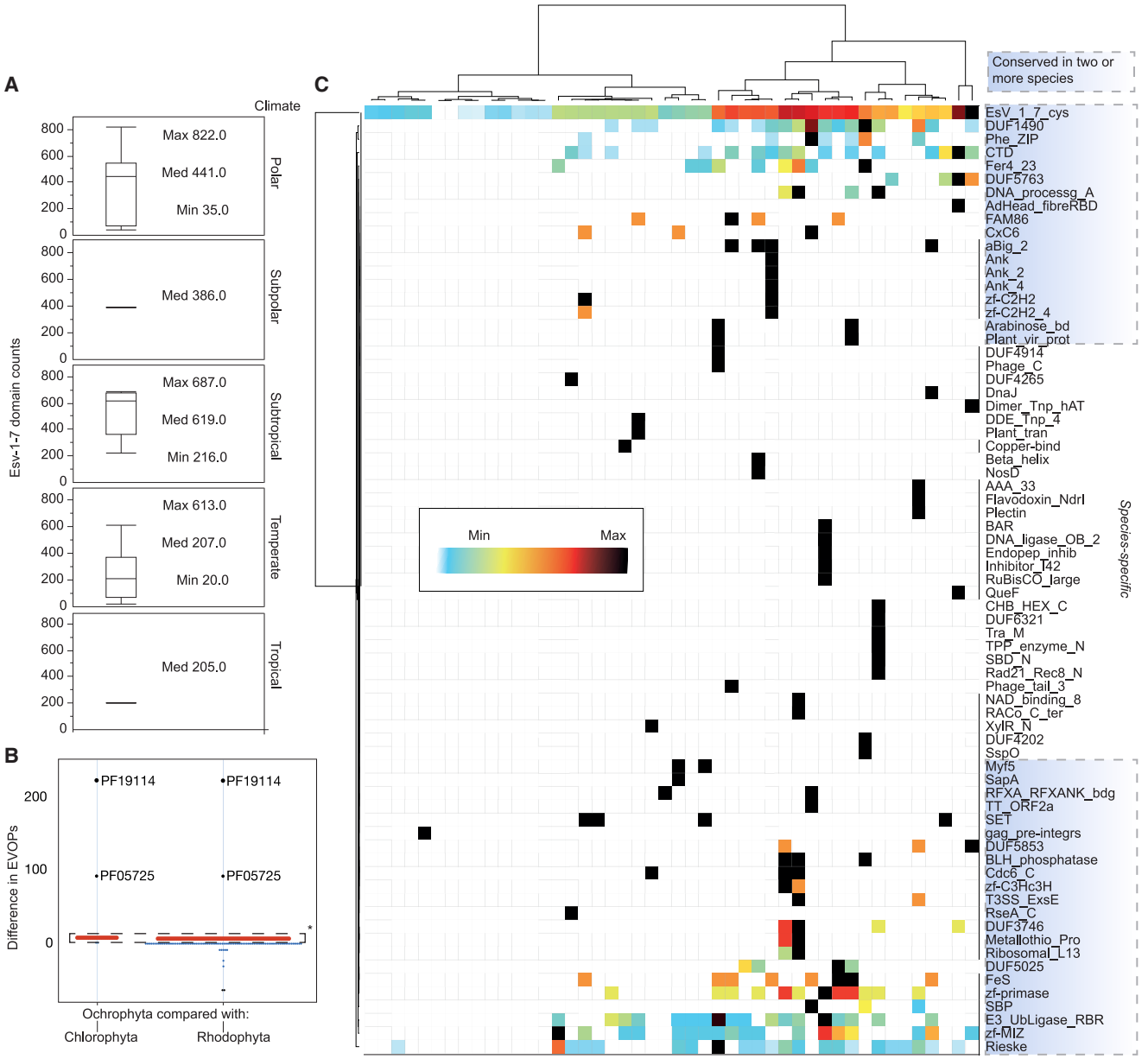


Figure 6. Phaeophyceae EsV-1-7 cysteine-rich motifs were found in macroalgal genes coding for a wide variety of functions (*n* = 43 genomes surveyed).

The FNIP repeat domain was also observed as a predominant feature in many Phaeophyceae genomes. Both motifs are common in viruses and amebae, and our data reveal that these sequences are conserved in Phaeophyceae as well. These motifs were found in a wide variety of genes with functions relating to molecular and physiological niche adaptation strategies, especially as they pertain to phaeophyte macroalgae.

(A) EsV-1-7 domain counts in phaeophyte macroalgae from different climate zones. A single line signifies that only one species was sampled in that zone. Our data show that ~300 EsV-1-7 domains could be recovered from each Phaeophyceae genome.

(B) Genes with VFAMs that also contained PFAMs (EVOP families) were compared between phyla (Supplemental Table 7). Phaeophyceae EVOPs were present at significantly higher levels than the other two clades, with FNIP and EsV-1-7 domains seen in drastically higher quantities than other EVOPs.

(C) Heatmap showing EsV-1-7 codomains. Remarkably, the multifunctional Rieske PFAM domains were common EsV-1-7 codomains.

Algal viruses can serve as metabolic gene reservoirs. For example, the Chlorovirus PBCV-1 has more than 400 protein-encoding genes, many of which are completely unexpected for a virus (Van Etten and Dunigan, 2012). Our past work with microalgal genomics revealed a vast network of viral gene transfers; here, we show that such donations may have generated scaffolds for complex multicellular developmental processes. The main contributions from viruses in each of the clades seem to be in redox balancing, transcriptional regulation, and metabolic flexibility. Further investigation of the role of viruses in the origin of subsequent adaptive radiations in especially the brown algae could provide valuable insights into a clade of growing environmental and industrial significance.

Please cite this article in press as: Nelson et al., Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity, Molecular Plant (2024), https://doi.org/10.1016/j.molp.2024.03.011

In summary, our work has successfully established a vast and varied genomic resource encompassing macroalgae from a wide range of phyla, climates, and ecological niches. Through meticulous comparative analyses, we have identified specific gene sets that elucidate both convergent and divergent evolutionary trajectories, contributing to the emergence of multicellularity and the rich morphological diversity observed in these organisms. We have highlighted critical genes and domains integral to processes such as cell adhesion, communication, polarity, and differentiation, which are fundamental in the development and sustenance of the complex multicellular structures in macroalgae. A significant emphasis of our study was on the role of endogenous viral elements and EVOPs, which play a pivotal role in the genomic diversification and innovative capacities of macroalgae. The datasets presented in this study not only deepen our understanding of the genetic underpinnings and environmental adaptability of macroalgae but also lay a robust foundation for future explorations and applications in the realm of genome-based macroalgal biology.

METHODS

Macroalgal cultures and sampling

The culture collection centers CCAP (www.ccap.ac.uk) and the National Center for Marine Algae and Microbiota (NCMA/CCMP; ncma.bigelow.org) provided the global macroalgal cultures; CCAP, 73 cultures and NCMA/CCMP, 39 cultures. This study has sequenced the genomes of all the available macroalgal cultures at the culture collection centers CCAP and CCMP between 2019 and 2021. The strains we used in this study covered a wide range of geographic locations, environments, and climatic zones (Supplemental Table 1). These included Rhodophyta (n = 65, excluding parasitic and unicellular species), Chlorophyta (n = 11, excluding higher plants, unicellular species, and seagrasses), and Ochrophyta macroalgae (n = 36, from the Ochrophyta phylum). These macroalgae were isolated from marine (n = 97), freshwater (n = 10), and brackish (n = 5) environments. Their native climates varied from polar, subpolar, temperate, subtropical, and tropical (Supplemental Table 1). Additionally, nine marine macroalgae (*Chondria dasyphylla*, rhodophyte; *Avrainvillea amadelpha*, chlorophyte; and *Padina boergesenii*, *Polycladia myrica*, *Sargassum latifolium*, *Sargassum angustifolium*, *Canistrocarpus cervicornis*, and two unidentified, phaeophytes) from the UAE (https://doi.org/10.5281/zenodo.7758508) were included in our analyses. All the strains were cultivated under 10 μmol photons/m^2/s at 13 h light:11 h dark cycle at NCMA; and under 20 μmol photons/m^2/s at 12 h light:12 h dark cycle at CCAP. The NCMA cultures were maintained in sterile Magenta GA-7-3 vessels with 0.5 L of medium (Supplemental Table 1). Cultures were axenically transferred every 12 weeks by dissecting a piece of macroalgal tissue and placing it in a new culture container of freshly filtered seawater with appropriate medium components aseptically added at the time of transfer.

DNA extraction, sequencing, and *de novo* genome assembly

Most DNA extractions were done using a phenol-chloroform protocol with the CTAB (cetyltrimethylammonium bromide; Sigma-Aldrich, St. Louis, MO, USA) extraction buffer (Phillips et al., 2001). The DNA from nine macroalgal culture samples (Supplemental Table 1) was extracted with the Dneasy Plant Pro Kit (Qiagen, Hilden, DE). We found that the traditional CTAB DNA extraction protocol yielded higher DNA concentrations and qualities. Sequencing of the 110 macroalgal genomes was performed with Illumina NexSeq 550, HiSeq 2500, and Novaseq PE150, aiming for ∼100× coverage in each assembly. We performed a benchmarking study using different assemblers. The results showed that CLC assembler (QIAGEN, 2022) yielded optimal assemblies with regard to Benchmarking Universal Single-Copy Orthologs (BUSCO, version 5.5.0) (Simao et al., 2015; Manni et al., 2021) scores and contig length (e.g., N50 values; Supplemental Figure 1). Hence, the CLC assembler was used for the 110 macroalgal assemblies here with three different k-mer sizes (40, 50, and 60). The best k-mer size for a given assembly was similarly evaluated, and the best assembly was used for downstream analyses. The macroalgal genomes with the highest N50s were of *Ulvella leptochaete* (chlorophyte, CCAP6037/1, N50 = 370 kbp) and *Chroothece richteriana* (rhodophyte, CCAP1353/4, N50 = 322 kbp). Phaeophyte macroalgae had higher average N50s (Supplemental Figure 1).

After decontamination, translated CDSs from the newly sequenced and reference genomes were analyzed with BUSCO to determine genome completeness using the latest (v5.5) BUSCO Docker (https://www.docker.com) container in Singularity (https://sylabs.io/) with the command ‘‘singularity run busco_v5.5.0_cv1.sif && singularity exec busco_v5.5.0_cv1.sif busco -c 16 -m protein –auto-lineage -i $LINE -o "$LINE".busco-out >> "$LINE"-busco-out’’ (also see Figure 1C–1E). The .fastq sequencing read files are available at NCBI (BioProject: PRJNA924561). The assemblies are available in Supplemental Data 1 (https://doi.org/10.5281/zenodo.7758534, https://zenodo.org/record/7758534). Briefly, our genomes showed similar size distributions as the published macroalgal genomes with Chlorophyta having the smallest range, then Rhodophyta, and Ochrophyta had the largest genomes. All ranges shown as interquartile (25%–75%) ranges: Chlorophyta, 90.11–267.46 Mbp, mean = 261.48 Mbp; Ochrophyta, 231.09–414.02 Mbp, mean = 320.27 Mbp; Rhodophyta, 143.28–348.50 Mbp, mean = 290.34 Mbp. The BUSCO metrics from the new genomes were higher, on average, than the published references both overall and within each phylum (Supplemental Table 1) (Chlorophyta, difference = 26.7 ± 18.5, p = 0.14; Ochrophyta, difference = 31.8 ± 12.5, p = 0.06; Rhodophyta, difference = 10.5 ± 5.3, p = 0.04) as analyzed with a standardized, Dockerized BUSCO (v5.5) instance.

ORF and contig decontamination

A small fraction of microbiome community members, endophytes, and contaminants were impossible to remove and were sequenced. After the unsatisfactory application of multiple available decontamination software, we developed a workflow to identify and remove contaminating CDSs and contigs. Contaminant sequences were thus removed from the macroalgal genome assemblies using the in-house decontamination pipeline, Basic Local Alignment Search Tool (BLAST)-Limiting Eukaryotic Assembly Contamination, Heuristically (BLEACH; see also

Please cite this article in press as: Nelson et al., Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity, Molecular Plant (2024), https://doi.org/10.1016/j.molp.2024.03.011

Supplemental Data 3). Assembled contigs from the sequenced macroalgal cultures were filtered according to this protocol to remove fungal, bacterial, and archaeal contamination. This approach relied on the homology of predicted CDSs to contaminants detected with Diamond BLAST (Buchfink et al., 2015, 2021). The end process removed as much contamination as possible and kept the macroalgal sequences by balancing two widely used, although heuristic, algorithms. Translated ORFs (tORFs) were queried against the full NCBI nr protein database with the command "diamond blastp –db nr –query $ –out $.diamond.blastp.txt –evalue $B –threads 28 -k 5 –outfmt 6 qseqid qlen sseqid sallseqid slen qstart qend sstart send qseq sseq evalue bitscore score length pident nident mismatch positive gapopen gaps ppos qframe btop stitle salltitles qcovhsp qtitle" for the first BLEACH round (i.e., default sensitivity). The second BLEACH round used the same BLASTP command as the first round except for the parameter "–more-sensitive" was used to further interrogate the sequences. The BLEACH rounds three to six further increased diamond BLAST sensitivity (e.g., most sensitive, ultra-sensitive) while increasing the allowed E value as a balancing mechanism to limit type I and II errors.

tORFs from the genomes were queried against the NCBI nr protein database using an accelerated version of the NCBI BLAST (Buchfink et al., 2015, 2021). Queries that had their best hit in a bacterial, fungal, or archaeal organism were removed in subsequent rounds with incrementally increased sensitivity. If a BLASTP search against nr revealed a bacterial hit, the corresponding CDS was heuristically labeled as a contaminant. Contigs sourcing contaminant CDSs were removed from the original assemblies. Reference macroalgal genomes, including *Ectocarpus siliculosus*, *Ectocarpus subulatus*, *Saccharina japonica*, *Chondrus crispus*, *Gracilariopsis chorda*, *Porphyra umbilicalis*, *Neopyropia yezoensis*, *Neoporphyra haitanensis*, *Gracilariopsis lemaneiformis*, *Kappaphycus alvarezii*, *Caulerpa lentillifera*, *Ulva mutabilis*, *Ulva prolifera*, and *Chara braunii* had few contaminant sequences (Supplemental Figure 2). After decontamination, the newly sequenced genomes had similar PFAM distributions to the reference genomes. After six rounds of BLEACH decontamination at incrementally increasing sensitivity, the cleaned genomes were used for downstream analyses (see also Supplemental Figure 2).

To control for over-cleaning, we examined the effect of BLEACH on assemblies that were not contaminated, including *Ectocarpus subulatus* and *Chlamydomonas reinhardtii* CC-1883 (Supplemental Data 3). The BLEACH process did not adversely affect these genomes, as 99.65% and 98.87% of contigs in each were retained. To validate our contig removal process, we mapped our original sequencing reads to the contaminant ORFs, extracted aligned reads, and mapped these reads to the clean contigs. Most reads from each assembly were marked as contaminants, and only 2.72% of the extracted reads mapped back to the clean contigs, indicating low rates of contaminants in the resultant filtered macroalgal genomes. We validated our decontamination pipeline by downsampling each tORFeome to 2000 sequences and performing NCBI BLASTP with the same parameters as the Diamond BLASTP. Estimations of percentage contamination are shown in Supplemental Figure 2 and listed in full in Supplemental Data 2 for each iteration of the BLEACH program.

Phylogenies and molecular-clock divergence estimates

Phylogenies of the sequenced species (Figure 1A) were reconstructed using PhyloT (https://phylot.biobyte.de/index.cgi) and the Interactive Tree of Life (https://itol.embl.de). The macroalgal species names from Supplemental Table 1 were used as queries to build the model based on established NCBI taxonomy.

Evolutionary timelines were reconstructed by averaging divergence estimates from multiple molecular-clock analyses (Hedges et al., 2004; Yoon et al., 2004; Zimmer et al., 2007; Parfrey et al., 2011; Yang et al., 2016) in timetree.org (Supplemental Figure 1). This software uses an extensive public database that unites thousands of published studies to form an interactive, time-scaled Tree of Life. The chronologies shown in Supplemental Figure 1 showcase pivotal moments in geological and atmospheric history for seamless comparison with evolutionary timelines and time-based trees of the relevant algal species.

Comparative genomics references

To discern macroalgal-specific gene collections in the three lineages, we performed a variety of comparative genomics using published reference genomes as validation standards. Decontaminated genomes and their translated ORFeomes were compared across phyla, climate, and habitats. Reference genomes included in our analyses were the phaeophyte macroalgae *Ectocarpus siliculosus* (model organism, ~196 Mbp) (Cock et al., 2010b), *Ectocarpus subulatus* (242 Mbp) (Dittami et al., 2020), and *Saccharina japonica* (commercial species, ~535 Mbp) (Ye et al., 2015); the rhodophyte *Chondrus crispus* (model organism, carrageenan production, ~104.98 Mbp) (Collen et al., 2013), *Gracilariopsis chorda* (92.18 Mbp) (Lee et al., 2018), *Porphyra umbilicalis* (commercial species, ~87.89 Mbp) (Brawley et al., 2017), *Neopyropia yezoensis* (commercially profitable, susabi-nori, ~107.59 Mb (Nakamura et al., 2013), *Neoporphyra haitanensis* (53.25 Mbp) (Cao et al., 2020), *Gracilariopsis lemaneiformis* (88.69 Mbp) (Zhou et al., 2013), and *Kappaphycus alvarezii* (commercial species, 336.72 Mbp) (Jia et al., 2020); and the Chlorophyta *Caulerpa lentillifera* (28.77 Mbp) (Arimoto et al., 2019), *Ulva mutabilis* (98.48 Mbp) (De Clerck et al., 2018), *Ulva prolifera* (887.88 Mbp) (He et al., 2021), and *Chara braunii* (1751.21 Mbp) (Nishiyama et al., 2018).

Structural and functional annotation

We used hidden Markov models to identify CDSs (https://github.com/KorfLab/SNAP), PFAMs (http://hmmer.org/), and VFAMs (https://derisilab.ucsf.edu/software/vFam/) from the macroalgal genomes. A custom *Arabidopsis thaliana* HMM yielded the most and highest-quality gene predictions for most species. To remove bias from variable gene calling due to other poor-quality HMMs, a "one-size-fits-all" approach was used with the *A. thaliana* HMM to predict CDS and translated peptides in the 124 macroalgal genomes. The newly sequenced genomes were predicted to be diploid or polyploid. In general, assemblies were haploid unless considerable heterozygosity was present.

The PFAM-A-v31.1 database was used as a query for translated ORFs. The command was "hmmsearch –noali -E 0.0001 –cpu 28 –domtblout $OUT $IN.aa.fa". All reported values in the columns in the output files are described in the HMMer user guide (http://eddylab.org/software/hmmer/Userguide.pdf). These include

Please cite this article in press as: Nelson et al., Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity, Molecular Plant (2024), https://doi.org/10.1016/j.molp.2024.03.011

sequence coordinates for alignment and matches with HMM models, E values, percentage identity, and bias. BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to find protein homology with the accelerated Diamond BLAST algorithm (Buchfink et al., 2015, 2021) for annotation and downstream decontamination. The Diamond BLASTP command was "diamond blastp –db nr –query $assembly.aa.fa –out $assembly.diamond.blastp.txt –more-sensitive –threads 4 -k 1 –outfmt 6 qseqid qlen sseqid sallseqid slen qstart qend sstart send qseq sseq evalue bitscore score length pident nident mismatch positive gapopen gaps ppos qframe btop stitle salltitles qcovhsp qtitle."

Enzyme Commission (EC) codes were obtained using PFAM associations described using the EC domain-miner tool (Alborzi et al., 2017). The association list is a downloadable file, and EC designations were obtained with the command: "while read l; do grep -w $l $PFAM.txt > $EC.out." Kyoto Encyclopedia of Genes and Genomes (KEGG) designations were obtained from EC IDs using EC2KEGG (Porollo, 2014). The EC codes were used to retrieve KEGG terms using EC2KEGG (Porollo, 2014). Briefly, the scripts in this package use "libwww-perl" (for Internet communication with KEGG, $http://search.cpan.org/$mschilli/libwww-perl-6.06/lib/LWP.pm), "Text-NSP" (for computing the Fisher exact test, $http://search.cpan.org/$tpederse/Text-NSP-1.27/), and "Statistics-Multtest" (for correcting p values on multiple hypotheses testing, $http://search.cpan.org/$jokergoo/Statistics-Multtest-0.13/). Nodes in the iPath diagram (Figure 3) represent enzymatic reactions with corresponding EC designations; bold edges indicate present ECs. For the iPATH analysis, EC codes were used to reconstruct metabolic pathways on the iPath website (https://pathways.embl.de/). The population of nodes in the reaction maps tightly corresponds with estimated origin times for each macroalgal lineage and further highlights the evolutionary innovations that multicellular algae have made in their respective habitats and genetic constraints.

Ternary analysis

To elucidate the propensity for PFAM domains to be found in the Rhodophyta, Chlorophyta, or Ochrophyta macroalgal genomes we surveyed, a ternary analysis was performed. Ternary analyses are commonly used in the physical sciences to determine the composition of components from three entities (see https://mathworld.wolfram.com/TernaryDiagram.html). Here, the three macroalgal lineages were interrogated regarding their ternary distribution to yield a high-level picture of the degree of unique and shared PFAMs. Ternary analysis was carried out in R using the package "ggtern" (http://www.ggtern.com), an extension to ggplot2. A normal simplex diagram graphs in k + 1 dimensions, whereas a ternary plot graphs in k = 2 with k + 1 = 3 vertices. Values for PFAM averages per phylum were used as inputs and represented either x, y, or z plots. The threshold of 80% purity was used to define clade-specific PFAMs, and PFAMs above this threshold in each of the phyla were used for downstream in domain-centric ontology enrichment analyses.

Domain-centric gene ontology enrichment

Functional enrichment analyses were carried out on subsets of PFAMs as indicated using Domain-centric Gene Ontology (dcGO; Fang and Gough, 2013). Preliminary HMMsearch (hmmer.org) was performed on translated ORFs with the command "hmmsearch –noali -E 0.0001 –cpu 28 –domtblout $OUT $IN.aa.fa." The software dcGO statistically infers associations between GO terms and combinations of PFAM domains. Uploaded queries were PFAM lists from selected sets in the ternary and Venn analyses. A false discovery rate (FDR) threshold of 0.01 was used for detection of enrichment and the full set of GO terms was queried. Briefly, distributions of GO terms are tested for overrepresentation (enrichment) against a collection of all GO terms. The analysis takes the hypergeometric distribution as a null-hypothesis statistical tests (equivalent to Fisher's exact tests) infers PFAM to GO associations. An overall p value and a corresponding overall hypergeometric score (*Z* score) is calculated by the observed PFAM count minus the expected PFAM count and divided by the standard deviation under the hypergeometric distribution with all available UniProt proteins. A secondary p value is calculated using only UniProt proteins in GO terms to establish a dual constraint filter for significance of matches. The strength of the inference is seen in the *Z* score.

Comparisons of micro- and macroalgal genomes

Microalgal tORFeomes were obtained from our previous work and HMMsearch was performed similarly to the macroalgal tORFeomes. Parsed HMMsearch results (i.e., PFAM matrices) were used in comparative analyses. Means were analyzed using one-way analysis of variance (ANOVA), where each comparison's p value, logworth, FDR (or adj. p value), and FDR logworth are reported (Supplemental Table 3). Protein family averages were compared across phyla between macroalgal and microalgal genome collections and screened for significant differences (Supplemental Table 4). The combined set included 251 algal genomes in total. The microalgal genome collection was composed of unicellular chlorophytes, ochrophytes, and rhodophytes sequenced in our previous work and the reference microalgal genomes included therein for analyses (Nelson et al., 2021). These included 76 chlorophytes, 44 ochrophytes, and seven rhodophytes. HMMsearch was applied on the microalgal tORFs in the same manner as the macroalgal genomes (Nelson et al., 2021). The combined micro-/macroalgal HMMsearch results were used comparative analyses using the entire PFAM set (~11 000 PFAMs) and reduced sets (e.g., the adhesome set of 110 PFAMs and the transporter set of 183 PFAMs).

Pathway analysis

Pathway discovery was facilitated by ECdomainMiner, a content-based filtering approach to automatically infer associations between EC codes and PFAM domains (Alborzi et al., 2017). Recovered EC codes were used to reconstruct metabolic pathways from the KEGG using iPATH3 (Letunic et al., 2008). The EC list can be uploaded to the iPATH3 website (https://pathways.embl.de) for an interactive exploration of the reactions and compounds comprising these macroalgal-specific pathways. Metabolic pathways were reconstructed using the indicated gene lists in the Interactive Pathways Explorer (iPath) (a web-based tool for the visualization, analysis, and customization of metabolic and other known pathway maps; https://pathways.embl.de/). EC codes were derived from PFAMs using the

Please cite this article in press as: Nelson et al., Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity, Molecular Plant (2024), https://doi.org/10.1016/j.molp.2024.03.011

ECDomainMiner software (Alborzi et al., 2017). This tool scavenges the protein data bank (PDB) (Bittrich et al., 2021; Flores et al., 2021; Behzadi and Gajdacs, 2022; Goodsell and Burley, 2022) to discover X-ray crystallography or NMR spectroscopy-resolved PDB entries corresponding to PFAM domains and classify them into EC codes. Only GOLD-ranked EC links were used.

Adhesome analysis

We defined the macroalgal adhesome as including cadherin, integrin, armadillo (ARM), lectin, F actin-binding CH, Kringle, PH, SH2, and LIM domains (n = 110 distinct domains). The accessions for the queried adhesome domains are in Supplemental Table 4. The cadherin PFAMs were queried were PF00028, PF01049, PF08266, PF08374, PF08758, PF12733, PF15974, PF16184, PF16492, PF17756, PF17803, PF17812, PF17813, PF17892, and PF18432. PFAM subsets were compared among phyla and native climate of the surveyed macroalgae (n Ochrophyta = 42, n Rhodophyta = 67, n Chlorophyta = 11). Means were analyzed using ANOVA, where each comparison's p value, logworth, FDR or adj. p value, and FDR logworth are reported in Supplemental Table 4.

Predictive modeling with neural networks

Predictive modeling using neural networks was carried out in the JMP Neural Networks module. Four sequential models were used to fit eight TanH, four linear, and one Gaussian formulae with a learning rate of 1.0. The input datasets were PFAM calls from the micro- and macroalgal genomes. Python, C, and Java code for the formulae necessary to reproduce the dNN are provided in the supplemental information. Deployment was on an high-performance computing (HPC) cluster with a LINUX CentOS operating system. Jobs were dispatched with SLURM; SLURM logs for programs run in this project are in the supplemental information for their corresponding analyses.

Viral family and EsV-1-7 (co)domain analyses

Hidden Markov models were used to identify VFAMs in the macroalgal ORFs as in Skewes-Cox et al. (2014) (https://derisilab.ucsf.edu/software/vFam/). Detected VFAMs were compared across phyla, habitat, and climate. As EsV-1-7 is a PFAM, not VFAM, domain, sequences containing EsV-1-7 domains were extracted from the PFAM HMMsearch results. Additional PFAMs on the EsV-1-7 sequences (i.e., codomains) were extracted in the same manner from the HMMsearch results (Supplemental Data 3, PFAM HMMSEARCH) and used for statistical comparative analyses by phyla, habitat, and climate.

Statistical analyses

One-way ANOVA statistical analyses (Blanca et al., 2017) were performed for comparisons of PFAMs, VFAMs, and their subsets. Batch *t-test*s comparing VFAM, PFAM, GO, and EVOP subsets were performed in the JMP Response Screening platform to screen for significant differences with practical effect sizes. See Supplemental Tables 1–5 in Supplemental Data 4 for raw and processed data frames used in these analyses. Standardized residues were used to compare means in >100 000 *t-test*s while controlling for FDR (Benjamini et al., 2001). Robust Huber M-estimation (using maximum-likelihood type estimators) (El Karoui et al., 2013) was used to reduce the sensitivity of the tests to outliers. Briefly, this method reduces the influence of outliers by minimizing their weights for the comparative statistics algorithm. The most informative information from these statistical tests for the general reader is in the "Response Screen Compare Means" tables, which include tests for all possible comparisons and results for practical equivalences and differences. Comparisons not passing either of these tests are labeled as "inconclusive." The practical difference is the difference in means having practical interest. Standard deviation estimates were computed from their interquartile ranges. The estimate was $\sigma = (IQR)/(1.3489795)$. The practical difference was computed as 6(σ) multiplied by 0.10 as a proxy for the practical difference proportion. Practical difference p values are given for tests of whether the absolute value of the mean difference in Y between comparison levels is less than or equal to the practical difference. Low p values indicate that the absolute difference exceeds the practical difference, indicating that the difference in the comparison is of practical significance.

Practical equivalence p values were computed using the two one-sided tests method to test for practical differences between means (Schuirmann, 1987). The practical difference specifies a threshold difference for which smaller differences are considered practically equivalent. One-sided *t-test*s are constructed for two null hypotheses: the true difference exceeds the practical difference; the true difference is less than the negative of the practical difference. If both tests reject, this indicates that the absolute difference in the means falls within the practical difference. Therefore, the groups are considered practically equivalent. The practical equivalence p value is the largest p value obtained on the one-sided *t-test*s. Low practical equivalence p values indicate that the mean response for the top comparison level is equivalent, in a practical sense, to the mean for the lower level.

DATA AND CODE AVAILABILITY

The .fastq sequencing read files are available at the NCBI sequence read archive (BioProject: PRJNA924561). The data and code used in data processing and analyses are available online in Zenodo repositories as described in the supplemental information: i.e., https://doi.org/10.5281/zenodo.7758534 (Supplemental Data 1, https://zenodo.org/record/7758534), https://doi.org/10.5281/zenodo.7751045 (Supplemental Data 2, https://zenodo.org/record/7751045), https://doi.org/10.5281/zenodo.7947301 (Supplemental Data 3, https://zenodo.org/record/7947301), and https://doi.org/10.5281/zenodo.10581453 (Supplemental Data 4, https://zenodo.org/records/10581453).

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Molecular Plant Online*.

FUNDING

This research was supported by NYUAD Faculty Research Funds (AD060) and Tamkeen under the NYU Abu Dhabi Research Institute Award to the NYUAD Center for Genomics and Systems Biology (73 71210 CGSB9). J.-C.T. is a Maitre de Recherche of the FRS-FNRS, Belgium. C.R.-M., F.D.B., and D.H.G. are employed at the Culture Collection of Algae and Protozoa, Scottish Association for Marine Science, and M.J.P. and M.W.L. are

employed at the National Center for Marine Algae and Microbiota, Bigelow Laboratory for Ocean Sciences.

AUTHOR CONTRIBUTIONS

K.S.-A. and A.M. conceptualized the study. C.R.-M., F.D.B., D.H.G., M.J.P., and M.W.L. cultured macroalgae and extracted DNA. K.S.-A., A.K.J., and D.R.N. designed computational data analysis. A.K.J., S.D., A.M., D.C.E.A., K.S.-A., and D.R.N. performed computational data analysis. A.M. and D.R.N. prepared the figures. W.C.R., A.M., J.-C.T., K.S.-A., and D.R.N. wrote the manuscript with input from all authors. K.S.-A. supervised the study.

ACKNOWLEDGMENTS

This research was carried out in part on the High Performance Computing resources at NYU and NYUAD. We would also like to thank Mehar Sultana, the NYUAD sequencing core of the Core Technologies Platform (CTP), Nizar Drou, and the NYUAD Bioinformatics core.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the authors used Open AI's Chat GPT4 to improve the readability of the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of this publication.

Received: October 6, 2023
Revised: January 31, 2024
Accepted: March 8, 2024
Published: April 12, 2024

REFERENCES

Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., Brown, M.W., Burki, F., Dunthorn, M., Hampl, V., et al. (2012). The revised classification of eukaryotes. J. Eukaryot. Microbiol. **59**:429–493. https://doi.org/10.1111/j.1550-7408.2012.00644.x.

Adl, S.M., Bass, D., Lane, C.E., Lukeš, J., Schoch, C.L., Smirnov, A., Agatha, S., Berney, C., Brown, M.W., Burki, F., et al. (2019). Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. J. Eukaryot. Microbiol. **66**:4–119. https://doi.org/10.1111/jeu.12691.

Alborzi, S.Z., Devignes, M.D., and Ritchie, D.W. (2017). ECDomainMiner: discovering hidden associations between enzyme commission numbers and Pfam domains. BMC Bioinf. **18**:107. https://doi.org/10.1186/s12859-017-1519-x.

Amack, J.D., and Manning, M.L. (2012). Knowing the boundaries: extending the differential adhesion hypothesis in embryonic cell sorting. Science **338**:212–215. https://doi.org/10.1126/science.1223953.

Aravind, L., and Koonin, E.V. (1999). Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. J. Mol. Biol. **287**:1023–1040. https://doi.org/10.1006/jmbi.1999.2653.

Arellano-Verdejo, J., and Lazcano-Hernández, H.E. (2021). Collective view: mapping Sargassum distribution along beaches. PeerJ. Comput. Sci. **7**:e528. https://doi.org/10.7717/peerj-cs.528.

Arimoto, A., Nishitsuji, K., Higa, Y., Arakaki, N., Hisata, K., Shinzato, C., Satoh, N., and Shoguchi, E. (2019). A siphonous macroalgal genome suggests convergent functions of homeobox genes in algae and land plants. DNA Res. **26**:183–192. https://doi.org/10.1093/dnares/dsz002.

Arun, A., Coelho, S.M., Peters, A.F., Bourdareau, S., Pérès, L., Scornet, D., Strittmatter, M., Lipinska, A.P., Yao, H., Godfroy, O., et al. (2019). Convergent recruitment of TALE homeodomain life cycle regulators to direct sporophyte development in land plants and brown algae. Elife **8**:e43101. https://doi.org/10.7554/eLife.43101.

Astrahan, P., Korzen, L., Khanin, M., Sharoni, Y., and Israel, Á. (2021). Seaweeds fast EDC bioremediation: Supporting evidence of EE2 and BPA degradation by the red seaweed Gracilaria sp., and a proposed model for the remedy of marine-borne phenol pollutants. Environ. Pollut. **278**:116853. https://doi.org/10.1016/j.envpol.2021.116853.

Balasubramanian, R.N., Gao, M., and Umen, J. (2023). Identification of cell-type specific alternative transcripts in the multicellular alga Volvox carteri. BMC Genom. **24**:654. https://doi.org/10.1186/s12864-023-09558-0.

Behzadi, P., and Gajdács, M. (2022). Worldwide Protein Data Bank (wwPDB): A virtual treasure for research in biotechnology. Eur. J. Microbiol. Immunol. **11**:77–86. https://doi.org/10.1556/1886.2021.00020.

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., and Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. Behav. Brain Res. **125**:279–284. https://doi.org/10.1016/s0166-4328(01)00297-2.

Berney, C., and Pawlowski, J. (2006). A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. Proc. Biol. Sci. **273**:1867–1872. https://doi.org/10.1098/rspb.2006.3537.

Bittrich, S., Rose, Y., Segura, J., Lowe, R., Westbrook, J.D., Duarte, J.M., and Burley, S.K. (2021). RCSB Protein Data Bank: Improved Annotation, Search, and Visualization of Membrane Protein Structures Archived in the PDB. Bioinformatics **38**:1452–1454. https://doi.org/10.1093/bioinformatics/btab813.

Blackstone, N.W., and Ellison, A.M. (2000). Maximal indirect development, set-aside cells, and levels of selection. J. Exp. Zool. **288**:99–104. https://doi.org/10.1002/1097-010x(20000815)288:2<99::aid-jez2>3.0.co;2-r.

Blanca, M.J., Alarcón, R., Arnau, J., Bono, R., and Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? Psicothema **29**:552–557. https://doi.org/10.7334/psicothema2016.383.

Blank, C.E. (2013). Origin and early evolution of photosynthetic eukaryotes in freshwater environments: reinterpreting proterozoic paleobiology and biogeochemical processes in light of trait evolution. J. Phycol. **49**:1040–1055. https://doi.org/10.1111/jpy.12111.

Bork, P., Doerks, T., Springer, T.A., and Snel, B. (1999). Domains in plexins: links to integrins and transcription factors. Trends Biochem. Sci. **24**:261–263. https://doi.org/10.1016/s0968-0004(99)01416-4.

Brawley, S.H., Blouin, N.A., Ficko-Blean, E., Wheeler, G.L., Lohr, M., Goodson, H.V., Jenkins, J.W., Blaby-Haas, C.E., Helliwell, K.E., Chan, C.X., et al. (2017). Insights into the red algae and eukaryotic evolution from the genome of Porphyra umbilicalis (Bangiophyceae, Rhodophyta). Proc. Natl. Acad. Sci. USA **114**:E6361–E6370. https://doi.org/10.1073/pnas.1703088114.

Bringloe, T.T., Starko, S., Wade, R.M., Vieira, C., Kawai, H., De Clerck, O., Cock, J.M., Coelho, S.M., Destombe, C., Valero, M., et al. (2020). Phylogeny and Evolution of the Brown Algae. Crit. Rev. Plant Sci. **39**:281–321. https://doi.org/10.1080/07352689.2020.1787679.

Broussard, J.A., and Green, K.J. (2022). Plectin pulls it together, coupling the cortical actin and intermediate filament cytoskeletons. J. Cell Biol. **221**:e202201054. https://doi.org/10.1083/jcb.202201054.

Brunet, T., and King, N. (2017). The Origin of Animal Multicellularity and Cell Differentiation. Dev. Cell **43**:124–140. https://doi.org/10.1016/j.devcel.2017.09.016.

Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods **12**:59–60. https://doi.org/10.1038/nmeth.3176.

Buchfink, B., Reuter, K., and Drost, H.G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat. Methods **18**:366–368. https://doi.org/10.1038/s41592-021-01101-x.

Campbell, I.D., and Humphries, M.J. (2011). Integrin structure, activation, and interactions. Cold Spring Harbor Perspect. Biol. **3**, a004994. https://doi.org/10.1101/cshperspect.a004994.

Cao, M., Xu, K., Yu, X., Bi, G., Liu, Y., Kong, F., Sun, P., Tang, X., Du, G., Ge, Y., et al. (2020). A chromosome-level genome assembly of *Pyropia haitanensis* (Bangiales, Rhodophyta). Mol. Ecol. Resour. **20**:216–227. https://doi.org/10.1111/1755-0998.13102.

Caspermeyer, J. (2017). Next Generation TimeTree: An Expanded History of Life on Earth at Your Fingertips. Mol. Biol. Evol. **34**:1822–1823. https://doi.org/10.1093/molbev/msx133.

Chae, H.B., Bae, S.B., Paeng, S.K., Wi, S.D., Phan, K.A.T., Kim, M.G., Kim, W.Y., Yun, D.J., and Lee, S.Y. (2023). The physiological role of thiol-based redox sensors in plant defense signaling. New Phytol. **239**:1203–1211. https://doi.org/10.1111/nph.19018.

Charrier, B., Abreu, M.H., Araujo, R., Bruhn, A., Coates, J.C., De Clerck, O., Katsaros, C., Robaina, R.R., and Wichard, T. (2017). Furthering knowledge of seaweed growth and development to facilitate sustainable aquaculture. New Phytol. **216**:967–975. https://doi.org/10.1111/nph.14728.

Cock, J.M., Coelho, S.M., Brownlee, C., and Taylor, A.R. (2010a). The Ectocarpus genome sequence: insights into brown algal biology and the evolutionary diversity of the eukaryotes. New Phytol. **188**:1–4. https://doi.org/10.1111/j.1469-8137.2010.03454.x.

Cock, J.M., Sterck, L., Rouzé, P., Scornet, D., Allen, A.E., Amoutzias, G., Anthouard, V., Artiguenave, F., Aury, J.M., Badger, J.H., et al. (2010b). The Ectocarpus genome and the independent evolution of multicellularity in brown algae. Nature **465**:617–621. https://doi.org/10.1038/nature09016.

Collén, J., Porcel, B., Carré, W., Ball, S.G., Chaparro, C., Tonon, T., Barbeyron, T., Michel, G., Noel, B., Valentin, K., et al. (2013). Genome structure and metabolic features in the red seaweed Chondrus crispus shed light on evolution of the Archaeplastida. Proc. Natl. Acad. Sci. USA **110**:5247–5252. https://doi.org/10.1073/pnas.1221259110.

Collins, C., Denisin, A.K., Pruitt, B.L., and Nelson, W.J. (2017). Changes in E-cadherin rigidity sensing regulate cell adhesion. Proc. Natl. Acad. Sci. USA **114**:E5835–E5844. https://doi.org/10.1073/pnas.1618676114.

Czimmerer, Z., and Nagy, L. (2023). Epigenomic regulation of macrophage polarization: Where do the nuclear receptors belong? Immunol. Rev. **317**:152–165. https://doi.org/10.1111/imr.13209.

De Clerck, O., Kao, S.M., Bogaert, K.A., Blomme, J., Foflonker, F., Kwantes, M., Vancaester, E., Vanderstraeten, L., Aydogdu, E., Boesger, J., et al. (2018). Insights into the Evolution of Multicellularity from the Sea Lettuce Genome. Curr. Biol. **28**:2921–2933.e5. https://doi.org/10.1016/j.cub.2018.08.015.

Deniz, F., and Ersanli, E.T. (2018). An ecofriendly approach for bioremediation of contaminated water environment: Potential contribution of a coastal seaweed community to environmental improvement. Int. J. Phytoremediation **20**:256–263. https://doi.org/10.1080/15226514.2017.1374335.

Di Trani, J.M., Liu, Z., Whitesell, L., Brzezinski, P., Cowen, L.E., and Rubinstein, J.L. (2022). Rieske head domain dynamics and indazole-derivative inhibition of Candida albicans complex III. Structure **30**:129–138.e4. https://doi.org/10.1016/j.str.2021.08.006.

Dittami, S.M., Corre, E., Brillet-Guéguen, L., Lipinska, A.P., Pontoizeau, N., Aite, M., Avia, K., Caron, C., Cho, C.H., Collén, J., et al. (2020). The genome of Ectocarpus subulatus - A highly stress-tolerant brown alga. Mar. Genomics **52**:100740. https://doi.org/10.1016/j.margen.2020.100740.

Domozych, D.S., and Domozych, C.E. (2014). Multicellularity in green algae: upsizing in a walled complex. Front. Plant Sci. **5**:649. https://doi.org/10.3389/fpls.2014.00649.

Dong, D., Hao, Q., Zhang, P., Wang, T., Han, F., Liang, X., and Fei, Z. (2019). Endoplasmic reticulum Ca2+ release causes Rieske iron-sulfur protein-mediated mitochondrial ROS generation in pulmonary artery smooth muscle cells. Biosci. Rep. **39**. https://doi.org/10.1042/BSR20192414.

Douzery, E.J.P., Snell, E.A., Bapteste, E., Delsuc, F., and Philippe, H. (2004). The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? Proc. Natl. Acad. Sci. USA **101**:15386–15391. https://doi.org/10.1073/pnas.0403984101.

El Karoui, N., Bean, D., Bickel, P.J., Lim, C., and Yu, B. (2013). On robust regression with high-dimensional predictors. Proc. Natl. Acad. Sci. USA **110**:14557–14562. https://doi.org/10.1073/pnas.1307842110.

Elizondo-González, R., Quiroz-Guzmán, E., Escobedo-Fregoso, C., Magallón-Servín, P., and Peña-Rodríguez, A. (2018). Use of seaweed Ulva lactuca for water bioremediation and as feed additive for white shrimp Litopenaeus vannamei. PeerJ **6**:e4459. https://doi.org/10.7717/peerj.4459.

Emily, M., and François, O. (2007). A statistical approach to estimating the strength of cell-cell interactions under the differential adhesion hypothesis. Theor. Biol. Med. Model. **4**:37. https://doi.org/10.1186/1742-4682-4-37.

Fang, H., and Gough, J. (2013). A domain-centric solution to functional genomics via dcGO Predictor. BMC Bioinf. **14** (*Suppl 3*):S9. https://doi.org/10.1186/1471-2105-14-S3-S9.

Fang, L., Leliaert, F., Novis, P.M., Zhang, Z., Zhu, H., Liu, G., Penny, D., and Zhong, B. (2018). Improving phylogenetic inference of core Chlorophyta using chloroplast sequences with strong phylogenetic signals and heterogeneous models. Mol. Phylogenet. Evol. **127**:248–255. https://doi.org/10.1016/j.ympev.2018.06.006.

Ferris, P., Olson, B.J.S.C., De Hoff, P.L., Douglass, S., Casero, D., Prochnik, S., Geng, S., Rai, R., Grimwood, J., Schmutz, J., et al. (2010). Evolution of an expanded sex-determining locus in Volvox. Science **328**:351–354. https://doi.org/10.1126/science.1186222.

Feyza Özgen, F., Runda, M.E., Burek, B.O., Wied, P., Bloh, J.Z., Kourist, R., and Schmidt, S. (2020). Artificial Light-Harvesting Complexes Enable Rieske Oxygenase Catalyzed Hydroxylations in Non-Photosynthetic cells. Angew. Chem., Int. Ed. Engl. **59**:3982–3987. https://doi.org/10.1002/anie.201914519.

Fiz-Palacios, O., Schneider, H., Heinrichs, J., and Savolainen, V. (2011). Diversification of land plants: insights from a family-level phylogenetic analysis. BMC Evol. Biol. **11**:341. https://doi.org/10.1186/1471-2148-11-341.

Flores, S.C., Alexiou, A., and Glaros, A. (2021). Mining the Protein Data Bank to improve prediction of changes in protein-protein binding. PLoS One **16**:e0257614. https://doi.org/10.1371/journal.pone.0257614.

Forzisi, E., and Sesti, F. (2022). Non-conducting functions of ion channels: The case of integrin-ion channel complexes. Channels **16**:185–197. https://doi.org/10.1080/19336950.2022.2108565.

Foty, R.A., and Steinberg, M.S. (2005). The differential adhesion hypothesis: a direct evaluation. Dev. Biol. **278**:255–263. https://doi.org/10.1016/j.ydbio.2004.11.012.

Fučíková, K.,L.F., Cooper, E.D., Škaloud, P., D'Hondt, S., De Clerck, O., Gurgel, C.F.D., Lewis, L.A., Lewis, P.O., Lopez-Bautista, J.M., Delwiche, C.F., et al. (2014). New phylogenetic hypotheses for the core Chlorophyta based on chloroplast sequence data. Front. Ecol. Evol. **2**:63. https://doi.org/10.3389/fevo.2014.00063.

Gaya, E., Fernández-Brime, S., Vargas, R., Lachlan, R.F., Gueidan, C., Ramírez-Mejía, M., and Lutzoni, F. (2015). The adaptive radiation of lichen-forming Teloschistaceae is associated with sunscreening pigments and a bark-to-rock substrate shift. Proc. Natl. Acad. Sci. USA **112**:11600–11605. https://doi.org/10.1073/pnas.1507072112.

Goff, L.J., and Coleman, A.W. (1987). The solution to the cytological paradox of isomorphy. J. Cell Biol. **104**:739–748. https://doi.org/10.1083/jcb.104.3.739.

Goodsell, D.S., and Burley, S.K. (2022). RCSB Protein Data Bank resources for structure-facilitated design of mRNA vaccines for existing and emerging viral pathogens. Structure **30**:55–68.e2. https://doi.org/10.1016/j.str.2021.10.008.

Goto, K., and Meyerowitz, E.M. (1994). Function and regulation of the Arabidopsis floral homeotic gene PISTILLATA. Genes Dev. **8**:1548–1560. https://doi.org/10.1101/gad.8.13.1548.

Gueidan, C., Ruibal, C., de Hoog, G.S., and Schneider, H. (2011). Rock-inhabiting fungi originated during periods of dry climate in the late Devonian and middle Triassic. Fungal Biol. **115**:987–996. https://doi.org/10.1016/j.funbio.2011.04.002.

He, P., Xu, S., Zhang, H., Wen, S., Dai, Y., Lin, S., and Yarish, C. (2008). Bioremediation efficiency in the removal of dissolved inorganic nutrients by the red seaweed, Porphyra yezoensis, cultivated in the open sea. Water Res. **42**:1281–1289. https://doi.org/10.1016/j.watres.2007.09.023.

He, Y., Shen, S., Yu, D., Wang, Y., Yin, J., Wang, Z., and Ye, Y. (2021). The *Ulva prolifera* genome reveals the mechanism of green tides. J. Oceanol. Limnol. **39**:1458–1470. https://doi.org/10.1007/s00343-020-0212-5.

Heckman, D.S., Geiser, D.M., Eidell, B.R., Stauffer, R.L., Kardos, N.L., and Hedges, S.B. (2001). Molecular evidence for the early colonization of land by fungi and plants. Science **293**:1129–1133. https://doi.org/10.1126/science.1061457.

Hedges, S.B., Blair, J.E., Venturi, M.L., and Shoe, J.L. (2004). A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evol. Biol. **4**:2. https://doi.org/10.1186/1471-2148-4-2.

Hernández-Vásquez, M.N., Adame-García, S.R., Hamoud, N., Chidiac, R., Reyes-Cruz, G., Gratton, J.P., Côté, J.F., and Vázquez-Prado, J. (2017). Cell adhesion controlled by adhesion G protein-coupled receptor GPR124/ADGRA2 is mediated by a protein complex comprising intersectins and Elmo-Dock. J. Biol. Chem. **292**:12178–12191. https://doi.org/10.1074/jbc.M117.780304.

Herron, M.D., Hackett, J.D., Aylward, F.O., and Michod, R.E. (2009). Triassic origin and early radiation of multicellular volvocine algae. Proc. Natl. Acad. Sci. USA **106**:3254–3258. https://doi.org/10.1073/pnas.0811205106.

Hess, S., Williams, S.K., Busch, A., Irisarri, I., Delwiche, C.F., de Vries, S., Darienko, T., Roger, A.J., Archibald, J.M., Buschmann, H., et al. (2022). A phylogenomically informed five-order system for the closest relatives of land plants. Curr. Biol. **32**:4473–4482.e7. https://doi.org/10.1016/j.cub.2022.08.022.

Hirano, K., Teraoka, T., Yamanaka, H., Harashima, A., Kunisaki, A., Takahashi, H., and Hosokawa, D. (2000). Novel mannose-binding rice lectin composed of some isolectins and its relation to a stress-inducible salT gene. Plant Cell Physiol. **41**:258–267. https://doi.org/10.1093/pcp/41.3.258.

Hou, Y.J., Guo, Y., Li, D.F., and Zhou, N.Y. (2021). Structural and Biochemical Analysis Reveals a Distinct Catalytic Site of Salicylate 5-Monooxygenase NagGH from Rieske Dioxygenases. Appl. Environ. Microbiol. **87**:e01629-20. https://doi.org/10.1128/AEM.01629-20.

Hou, Z., Ma, X., Shi, X., Li, X., Yang, L., Xiao, S., De Clerck, O., Leliaert, F., and Zhong, B. (2022). Phylotranscriptomic insights into a Mesoproterozoic-Neoproterozoic origin and early radiation of green seaweeds (Ulvophyceae). Nat. Commun. **13**:1610. https://doi.org/10.1038/s41467-022-29282-9.

Iranzo, J., Lobkovsky, A.E., Wolf, Y.I., and Koonin, E.V. (2014). Virus-host arms race at the joint origin of multicellularity and programmed cell death. Cell Cycle **13**:3083–3088. https://doi.org/10.4161/15384101.2014.949496.

Jia, S., Wang, G., Liu, G., Qu, J., Zhao, B., Jin, X., Zhang, L., Yin, J., Liu, C., Shan, G., et al. (2020). High-quality *de novo* genome assembly of *Kappaphycus alvarezii* based on both PacBio and HiSeq sequencing. Preprint at bioRxiv. https://doi.org/10.1101/2020.02.15.950402.

Katz, L., Sirjacobs, D., Gobert, S., Lejeune, P., and Danis, B. (2021). Distribution of macroalgae in the area of Calvi (Corsica). Biodivers. Data J. **9**:e68249. https://doi.org/10.3897/BDJ.9.e68249.

Kidgell, J.T., de Nys, R., Hu, Y., Paul, N.A., and Roberts, D.A. (2014). Bioremediation of a complex industrial effluent by biosorbents derived from freshwater macroalgae. PLoS One **9**:e94706. https://doi.org/10.1371/journal.pone.0094706.

Klein, J., Saedler, H., and Huijser, P. (1996). A new family of DNA binding proteins includes putative transcriptional regulators of the Antirrhinum majus floral meristem identity gene SQUAMOSA. Mol. Gen. Genet. **250**:7–16. https://doi.org/10.1007/BF02191820.

Kloareg, B., Badis, Y., Cock, J.M., and Michel, G. (2021). Role and Evolution of the Extracellular Matrix in the Acquisition of Complex Multicellularity in Eukaryotes: A Macroalgal Perspective. Genes **12**:1059. https://doi.org/10.3390/genes12071059.

Kropf, D.L. (1997). Induction of Polarity in Fucoid Zygotes. Plant Cell **9**:1011–1020. https://doi.org/10.1105/tpc.9.7.1011.

Lang, D., Weiche, B., Timmerhaus, G., Richardt, S., Riaño-Pachón, D.M., Corrêa, L.G.G., Reski, R., Mueller-Roeber, B., and Rensing, S.A. (2010). Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. Genome Biol. Evol. **2**:488–503. https://doi.org/10.1093/gbe/evq032.

Lee, J., Yang, E.C., Graf, L., Yang, J.H., Qiu, H., Zelzion, U., Chan, C.X., Stephens, T.G., Weber, A.P.M., Boo, G.H., et al. (2018). Analysis of the Draft Genome of the Red Seaweed *Gracilariopsis chorda* Provides Insights into Genome Size Evolution in Rhodophyta. Mol. Biol. Evol. **35**:1869–1886. https://doi.org/10.1093/molbev/msy081.

Letunic, I., Yamada, T., Kanehisa, M., and Bork, P. (2008). iPath: interactive exploration of biochemical pathways and networks. Trends Biochem. Sci. **33**:101–103. https://doi.org/10.1016/j.tibs.2008.01.001.

Littler, M.M., Littler, D.S., Blair, S.M., and NORRIS, J.N. (1985). Deepest Known Plant Life Discovered on an Uncharted Seamount. Science **227**:57–59. https://doi.org/10.1126/science.227.4682.57.

Liu, W., Xin, M.L., Zhou, J., Zhan, D.M., Ding, G., and Wu, H.Y. (2021). Habitat hierarchies distribution of *Sargassum muticum* in Lidao bay, Shandong, China based on habitat suitability index model. Yingyong Shengtai Xuebao **32**:1061–1068. https://doi.org/10.13287/j.1001-9332.202103.034.

Luo, H., Wang, Q., Liu, Z., Wang, S., Long, A., and Yang, Y. (2020). Potential bioremediation effects of seaweed Gracilaria lemaneiformis on heavy metals in coastal sediment from a typical mariculture zone. Chemosphere **245**:125636. https://doi.org/10.1016/j.chemosphere.2019.125636.

Ma, L., Wang, Y., Liu, W., and Liu, Z. (2014). Overexpression of an alfalfa GDP-mannose 3, 5-epimerase gene enhances acid, drought and salt tolerance in transgenic Arabidopsis by increasing ascorbate accumulation. Biotechnol. Lett. **36**:2331–2341. https://doi.org/10.1007/s10529-014-1598-y.

Macaisne, N., Liu, F., Scornet, D., Peters, A.F., Lipinska, A., Perrineau, M.M., Henry, A., Strittmatter, M., Coelho, S.M., and Cock, J.M. (2017). The Ectocarpus IMMEDIATE UPRIGHT gene encodes a member of a novel family of cysteine-rich proteins with an unusual distribution across the eukaryotes. Development **144**:409–418. https://doi.org/10.1242/dev.141523.

Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., and Zdobnov, E.M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol. Biol. Evol. **38**:4647–4654. https://doi.org/10.1093/molbev/msab199.

Mansour, A.T., Alprol, A.E., Ashour, M., Ramadan, K.M.A., Alhajji, A.H.M., and Abualnaja, K.M. (2022). Do Red Seaweed Nanoparticles Enhance Bioremediation Capacity of Toxic Dyes from Aqueous Solution? Gels **8**:310. https://doi.org/10.3390/gels8050310.

Marks, P.C., Hewitt, B.R., Baird, M.A., Wiche, G., and Petrie, R.J. (2022). Plectin linkages are mechanosensitive and required for the nuclear piston mechanism of three-dimensional cell migration. Mol. Biol. Cell **33**:ar104. https://doi.org/10.1091/mbc.E21-08-0414.

Matt, G., and Umen, J. (2016). Volvox: A simple algal model for embryogenesis, morphogenesis and cellular differentiation. Dev. Biol. **419**:99–113. https://doi.org/10.1016/j.ydbio.2016.07.014.

Matt, G.Y., and Umen, J.G. (2018). Cell-Type Transcriptomes of the Multicellular Green Alga Volvox carteri Yield Insights into the Evolutionary Origins of Germ and Somatic Differentiation Programs. G3 (Bethesda) **8**:531–550. https://doi.org/10.1534/g3.117.300253.

Mazéas, L., Yonamine, R., Barbeyron, T., Henrissat, B., Drula, E., Terrapon, N., Nagasato, C., and Hervé, C. (2022). Assembly and synthesis of the extracellular matrix in brown algae. Semin. Cell Dev. Biol. **134**:112–124. https://doi.org/10.1016/j.semcdb.2022.03.005.

McDonald, A.G., and Tipton, K.F. (2023). Enzyme nomenclature and classification: the state of the art. FEBS J. **290**:2214–2231. https://doi.org/10.1111/febs.16274.

Mei, L., Zheng, Y.M., Song, T., Yadav, V.R., Joseph, L.C., Truong, L., Kandhi, S., Barroso, M.M., Takeshima, H., Judson, M.A., and Wang, Y.X. (2020). Rieske iron-sulfur protein induces FKBP12.6/RyR2 complex remodeling and subsequent pulmonary hypertension through NF-kappaB/cyclin D1 pathway. Nat. Commun. **11**:3527. https://doi.org/10.1038/s41467-020-17314-1.

Mulders, Y., Filbee-Dexter, K., Bell, S., Bosch, N.E., Pessarrodona, A., Sahin, D., Vranken, S., Zarco-Perello, S., and Wernberg, T. (2022). Intergrading reef communities across discrete seaweed habitats in a temperate-tropical transition zone: Lessons for species reshuffling in a warming ocean. Ecol. Evol. **12**:e8538. https://doi.org/10.1002/ece3.8538.

Müller, D.G. (1966). [Studies on the life cycle of the brown alga Ectocarpus siliculosus from Naples, Italy]. Planta **68**:57–68. https://doi.org/10.1007/BF00385371.

Müller, D.G. (1967a). [A readily volatile gyno-gamone from the brown alga Ectocarpus siliculosus]. Naturwissenschaften **54**:496–497. https://doi.org/10.1007/BF00702523.

Müller, D.G. (1967b). [Culture experiments on life cycle, nuclear phases, and sexuality of the brown alga Ectocarpus siliculosus]. Planta **75**:39–54. https://doi.org/10.1007/BF00380838.

Müller, D.G. (1968). [Investigations on the nature of a chemotactic substance produced by the female gametes of Ectocarpus siliculosus : I. Methods, isolation, and detection by Gas Chromatography]. Planta **81**:160–168. https://doi.org/10.1007/BF00417445.

Munakata, H., Nakada, T., Nakahigashi, K., Nozaki, H., and Tomita, M. (2016). Phylogenetic Position and Molecular Chronology of a Colonial Green Flagellate, Stephanosphaera pluvialis (Volvocales, Chlorophyceae), among Unicellular Algae. J. Eukaryot. Microbiol. **63**:340–348. https://doi.org/10.1111/jeu.12283.

Nagy, L.G., Kovács, G.M., and Krizsán, K. (2018). Complex multicellularity in fungi: evolutionary convergence, single origin, or both? Biol. Rev. Camb. Phil. Soc. **93**:1778–1794. https://doi.org/10.1111/brv.12418.

Nakamura, Y., Sasaki, N., Kobayashi, M., Ojima, N., Yasuike, M., Shigenobu, Y., Satomi, M., Fukuma, Y., Shiwaku, K., Tsujimoto, A., et al. (2013). The first symbiont-free genome sequence of marine red alga, Susabi-nori (Pyropia yezoensis). PLoS One **8**:e57122. https://doi.org/10.1371/journal.pone.0057122.

Nelson, D.R., Hazzouri, K.M., Lauersen, K.J., Jaiswal, A., Chaiboonchoe, A., Mystikou, A., Fu, W., Daakour, S., Dohai, B., Alzahmi, A., et al. (2021). Large-scale genome sequencing reveals the driving forces of viruses in microalgal evolution. Cell Host Microbe **29**:250–266.e8. https://doi.org/10.1016/j.chom.2020.12.005.

Nicol, A., and Garrod, D.R. (1979). The sorting out of embryonic cells in monolayer, the differential adhesion hypothesis and the non-specificity of cell adhesion. J. Cell Sci. **38**:249–266. https://doi.org/10.1242/jcs.38.1.249.

Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K.K., Haas, F.B., Vanderstraeten, L., Becker, D., Lang, D., et al. (2018). The Chara Genome: Secondary Complexity and Implications for Plant Terrestrialization. Cell **174**:448–464.e24. https://doi.org/10.1016/j.cell.2018.06.033.

Nji Wandi, B., Siitonen, V., Palmu, K., and Metsä-Ketelä, M. (2020). The Rieske Oxygenase SnoT Catalyzes 2''-Hydroxylation of l-Rhodosamine in Nogalamycin Biosynthesis. Chembiochem **21**:3062–3066. https://doi.org/10.1002/cbic.202000229.

Ohme-Takagi, M., and Shinshi, H. (1995). Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. Plant Cell **7**:173–182. https://doi.org/10.1105/tpc.7.2.173.

Parfrey, L.W., Lahr, D.J.G., Knoll, A.H., and Katz, L.A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc. Natl. Acad. Sci. USA **108**:13624–13629. https://doi.org/10.1073/pnas.1110633108.

Conlin, P.L., Goldsby, H.J., Libby, E., Skocelas, K.G., Ratcliff, W.C., Ofria, C., and Kerr, B. (2023). Division of labor promotes the entrenchment of multicellularity. Preprint at bioRxiv. https://doi.org/10.1101/2023.03.15.532780.

Phillips, N., Smith, C.M., and Morden, C.W. (2001). An effective DNA extraction protocol for brown algae. Phycol. Res. **49**:97–102. https://doi.org/10.1046/j.1440-1835.2001.00229.x.

Ponsen, M., Tuyls, K., Kaisers, M., and Ramon, J. (2009). An evolutionary game-theoretic analysis of poker strategies. Entertainment Computing **1**:39–45. https://doi.org/10.1016/j.entcom.2009.09.002.

Porollo, A. (2014). EC2KEGG: a command line tool for comparison of metabolic pathways. Source Code Biol Med **9**:19. https://doi.org/10.1186/1751-0473-9-19.

Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., and Finn, R.D. (2018). HMMER web server: 2018 update. Nucleic Acids Res. **46**:W200–W204. https://doi.org/10.1093/nar/gky448.

Prechova, M., Adamova, Z., Schweizer, A.L., Maninova, M., Bauer, A., Kah, D., Meier-Menches, S.M., Wiche, G., Fabry, B., and Gregor, M. (2022). Plectin-mediated cytoskeletal crosstalk controls cell tension and cohesion in epithelial sheets. J. Cell Biol. **221**:e202105146. https://doi.org/10.1083/jcb.202105146.

Prochnik, S.E., Umen, J., Nedelcu, A.M., Hallmann, A., Miller, S.M., Nishii, I., Ferris, P., Kuo, A., Mitros, T., Fritz-Laylin, L.K., et al. (2010). Genomic analysis of organismal complexity in the multicellular green alga Volvox carteri. Science **329**:223–226. https://doi.org/10.1126/science.1188800.

Purhonen, J., Grigorjev, V., Ekiert, R., Aho, N., Rajendran, J., Pietras, R., Truvé, K., Wikström, M., Sharma, V., Osyczka, A., et al. (2020). A spontaneous mitonuclear epistasis converging on Rieske Fe-S protein exacerbates complex III deficiency in mice. Nat. Commun. **11**:322. https://doi.org/10.1038/s41467-019-14201-2.22.0

QIAGEN. (2022). In QIAGEN CLC Genomics Workbench 20.0, Q.C.G.W. 22.0, ed.

Ratcliff, W.C., Denison, R.F., Borrello, M., and Travisano, M. (2012). Experimental evolution of multicellularity. Proc. Natl. Acad. Sci. USA **109**:1595–1600. https://doi.org/10.1073/pnas.1115323109.

Roberts, D.A., Paul, N.A., Bird, M.I., and de Nys, R. (2015). Bioremediation for coal-fired power stations using macroalgae. J. Environ. Manag. **153**:25–32. https://doi.org/10.1016/j.jenvman.2015.01.036.

Schuirmann, D.J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J. Pharmacokinet. Biopharm. **15**:657–680. https://doi.org/10.1007/BF01068419.

Sharma, A., Wolfgruber, T.K., and Presting, G.G. (2013). Tandem repeats derived from centromeric retrotransposons. BMC Genom. **14**:142. https://doi.org/10.1186/1471-2164-14-142.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics **31**:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Skewes-Cox, P., Sharpton, T.J., Pollard, K.S., and DeRisi, J.L. (2014). Profile hidden Markov models for the detection of viruses within metagenomic sequence data. PLoS One **9**:e105067. https://doi.org/10.1371/journal.pone.0105067.

Sode, S., Bruhn, A., Balsby, T.J.S., Larsen, M.M., Gotfredsen, A., and Rasmussen, M.B. (2013). Bioremediation of reject water from anaerobically digested waste water sludge with macroalgae (Ulva lactuca, Chlorophyta). Bioresour. Technol. **146**:426–435. https://doi.org/10.1016/j.biortech.2013.06.062.

Song, X.H., Assis, J., Zhang, J., Gao, X., Gao, H.G., Duan, D.L., Serrão, E.A., and Hu, Z.M. (2021). Climate-induced range shifts shaped the present and threaten the future genetic variability of a marine brown alga in the Northwest Pacific. Evol. Appl. **14**:1867–1879. https://doi.org/10.1111/eva.13247.

Steinberg, M.S. (1975). Adhesion-guided multicellular assembly: a commentary upon the postulates, real and imagined, of the differential adhesion hypothesis, with special attention to computer simulations of cell sorting. J. Theor. Biol. **55**:431–443. https://doi.org/10.1016/s0022-5193(75)80091-9.

Stelling-Wood, T.P., Poore, A.G.B., and Gribben, P.E. (2021). Shifts in biomass and structure of habitat-formers across a latitudinal gradient. Ecol. Evol. **11**:8831–8842. https://doi.org/10.1002/ece3.7714.

Stengel, D.B., Macken, A., Morrison, L., and Morley, N. (2004). Zinc concentrations in marine macroalgae and a lichen from western Ireland in relation to phylogenetic grouping, habitat and morphology. Mar. Pollut. Bull. **48**:902–909. https://doi.org/10.1016/j.marpolbul.2003.11.014.

Strother, P.K., Brasier, M.D., Wacey, D., Timpe, L., Saunders, M., and Wellman, C.H. (2021). A possible billion-year-old holozoan with differentiated multicellularity. Curr. Biol. **31**:2658–2665.e2. https://doi.org/10.1016/j.cub.2021.03.051.

Sun, X., Liu, Z., Jiang, Q., and Yang, Y. (2019). Concentrations of various elements in seaweed and seawater from Shen'ao Bay, Nan'ao Island, Guangdong coast, China: Environmental monitoring and the bioremediation potential of the seaweed. Sci. Total Environ. **659**:632–639. https://doi.org/10.1016/j.scitotenv.2018.12.364.

Thomas, W.A., and Yancey, J. (1988). Can retinal adhesion mechanisms determine cell-sorting patterns: a test of the differential adhesion hypothesis. Development **103**:37–48. https://doi.org/10.1242/dev.103.1.37.

Thompson, C.J., Vu, V.H., Leckband, D.E., and Schwartz, D.K. (2021). Cadherin cis and trans interactions are mutually cooperative. Proc. Natl. Acad. Sci. USA **118**:e2019845118. https://doi.org/10.1073/pnas.2019845118.

Tong, C.F., Zhang, Y., Lü, S.Q., Li, N., Gong, Y.X., Yang, H., Feng, S.L., Du, Y., Huang, D.D., and Long, M. (2018). Binding of intercellular adhesion molecule 1 to beta2-integrin regulates distinct cell adhesion processes on hepatic and cerebral endothelium. Am. J. Physiol. Cell Physiol. **315**:C409–C421. https://doi.org/10.1152/ajpcell.00083.2017.

Ulmasov, T., Hagen, G., and Guilfoyle, T.J. (1997). ARF1, a transcription factor that binds to auxin response elements. Science **276**:1865–1868. https://doi.org/10.1126/science.276.5320.1865.

Umen, J.G. (2020). Volvox and volvocine green algae. EvoDevo **11**:13. https://doi.org/10.1186/s13227-020-00158-7.

Van Etten, J.L., and Dunigan, D.D. (2012). Chloroviruses: not your everyday plant virus. Trends Plant Sci. **17**:1–8. https://doi.org/10.1016/j.tplants.2011.10.005.

Venkatesan, J., Anil, S., Rao, S., Bhatnagar, I., and Kim, S.K. (2019). Sulfated Polysaccharides from Macroalgae for Bone Tissue Regeneration. Curr. Pharmaceut. Des. **25**:1200–1209. https://doi.org/10.2174/1381612825666190425161630.

Vreeland, V., Grotkopp, E., Espinosa, S., Quiroz, D., Laetsch, W.M., and West, J. (1993). The pattern of cell wall adhesive formation by Fucus zygotes. Hydrobiologia **260–261**:485–491. https://doi.org/10.1007/BF00049060.

Whittaker, C.A., Bergeron, K.F., Whittle, J., Brandhorst, B.P., Burke, R.D., and Hynes, R.O. (2006). The echinoderm adhesome. Dev. Biol. **300**:252–266. https://doi.org/10.1016/j.ydbio.2006.07.044.

Wiche, G. (2022). Plectin in Health and Disease. Cells **11**. https://doi.org/10.3390/cells11091412.

Winograd-Katz, S.E., Fässler, R., Geiger, B., and Legate, K.R. (2014). The integrin adhesome: from genes and proteins to human disease. Nat. Rev. Mol. Cell Biol. **15**:273–288. https://doi.org/10.1038/nrm3769.

Wiseman, L.L. (1977). Can the differential adhesion hypothesis explain how an aggregate of strongly cohesive cells can be penetrated by more weakly cohesive cells? Dev. Biol. **58**:204–211. https://doi.org/10.1016/0012-1606(77)90086-0.

Xia, Y., Cheng, H., Bian, W., Wang, W., Zhu, M., and Zhong, J. (2022). Role of an FNIP Repeat Domain-Containing Protein Encoded by Megavirus Baoshan during Viral Infection. J. Virol. **96**:e0081322. https://doi.org/10.1128/jvi.00813-22.

Yamamoto, K., Hamaji, T., Kawai-Toyooka, H., Matsuzaki, R., Takahashi, F., Nishimura, Y., Kawachi, M., Noguchi, H., Minakuchi, Y., Umen, J.G., et al. (2021). Three genomes in the algal genus Volvox reveal the fate of a haploid sex-determining region after a transition to homothallism. Proc. Natl. Acad. Sci. USA **118**. https://doi.org/10.1073/pnas.2100712118.

Yang, E.C., Boo, S.M., Bhattacharya, D., Saunders, G.W., Knoll, A.H., Fredericq, S., Graf, L., and Yoon, H.S. (2016). Divergence time estimates and the evolution of major lineages in the florideophyte red algae. Sci. Rep. **6**:21361. https://doi.org/10.1038/srep21361.

Ye, N., Zhang, X., Miao, M., Fan, X., Zheng, Y., Xu, D., Wang, J., Zhou, L., Wang, D., Gao, Y., et al. (2015). *Saccharina* genomes provide novel insight into kelp biology. Nat. Commun. **6**:6986. https://doi.org/10.1038/ncomms7986.

Yip, Z.T., Quek, R.Z.B., and Huang, D. (2020). Historical biogeography of the widespread macroalga Sargassum (Fucales, Phaeophyceae). J. Phycol. **56**:300–309. https://doi.org/10.1111/jpy.12945.

Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G., and Bhattacharya, D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. Mol. Biol. Evol. **21**:809–818. https://doi.org/10.1093/molbev/msh075.

Yoshiie, T., Maeda, M., Kimura, M., Hama, Y., Uchida, M., and Kimura, Y. (2012). Structural features of N-glycans of seaweed glycoproteins: predominant occurrence of high-mannose type N-glycans in marine plants. Biosci. Biotechnol. Biochem. **76**:1996–1998. https://doi.org/10.1271/bbb.120463.

Yu, C., Yan, M., Dong, H., Luo, J., Ke, Y., Guo, A., Chen, Y., Zhang, J., and Huang, X. (2021). Maize bHLH55 functions positively in salt tolerance through modulation of AsA biosynthesis by directly regulating GDP-mannose pathway genes. Plant Sci. **302**:110676. https://doi.org/10.1016/j.plantsci.2020.110676.

Zaidel-Bar, R. (2013). Cadherin adhesome at a glance. J. Cell Sci. **126**:373–378. https://doi.org/10.1242/jcs.111559.

Zaidel-Bar, R., and Geiger, B. (2010). The switchable integrin adhesome. J. Cell Sci. **123**:1385–1388. https://doi.org/10.1242/jcs.066183.

Zaidel-Bar, R., Itzkovitz, S., Ma'ayan, A., Iyengar, R., and Geiger, B. (2007). Functional atlas of the integrin adhesome. Nat. Cell Biol. **9**:858–867. https://doi.org/10.1038/ncb0807-858.

Zheng, Y., and Leftheris, K. (2020). Insights into Protein-Ligand Interactions in Integrin Complexes: Advances in Structure Determinations. J. Med. Chem. **63**:5675–5696. https://doi.org/10.1021/acs.jmedchem.9b01869.

Zhou, W., Hu, Y.Y., Sui, Z.H., Fu, F., Wang, J.G., Chang, L.P., Guo, W.H., and Li, B.B. (2013). Genome Survey Sequencing and Genetic Background Characterization of Gracilariopsis lemaneiformis (Rhodophyta) Based on Next-Generation Sequencing. PLoS One **8**. https://doi.org/10.1371/journal.pone.0069909.

Zimmer, A., Lang, D., Richardt, S., Frank, W., Reski, R., and Rensing, S.A. (2007). Dating the early evolution of plants: detection and molecular clock analyses of orthologs. Mol. Genet. Genomics. **278**:393–402. https://doi.org/10.1007/s00438-007-0257-6.

Znad, H., Awual, M.R., and Martini, S. (2022). The Utilization of Algae and Seaweed Biomass for Bioremediation of Heavy Metal-Contaminated Wastewater. Molecules **27**:1275. https://doi.org/10.3390/molecules27041275.

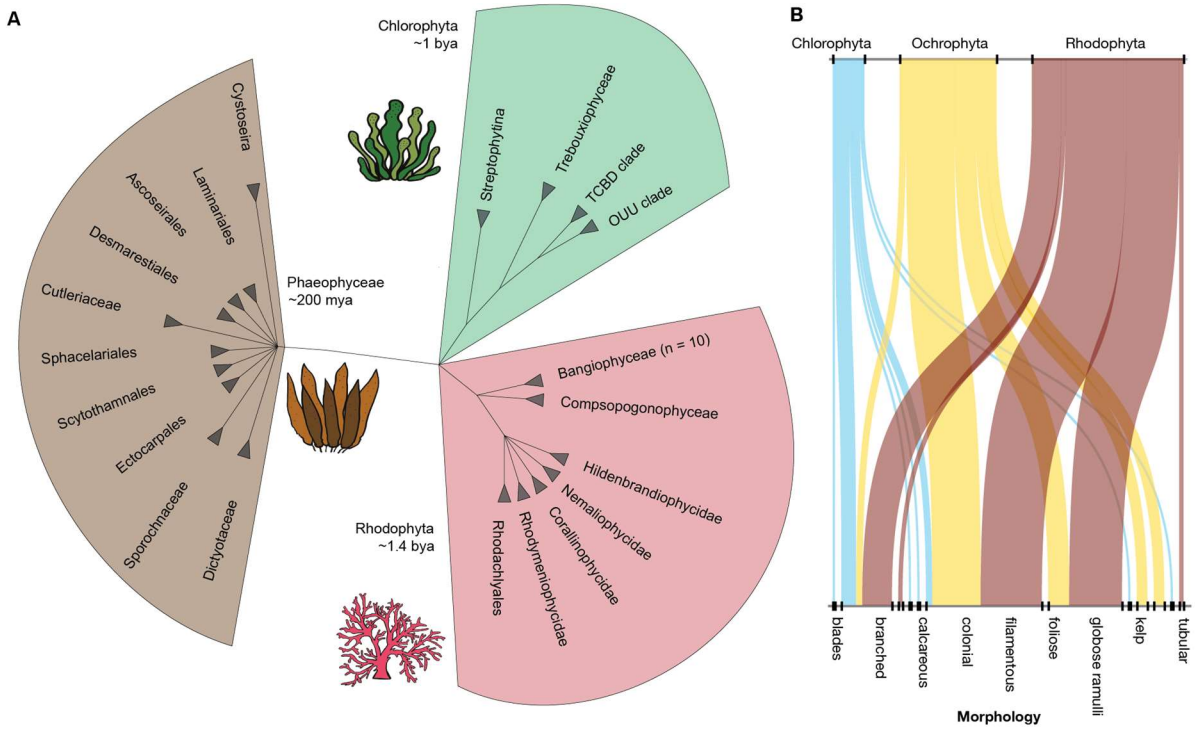
Molecular Plant, Volume *17*

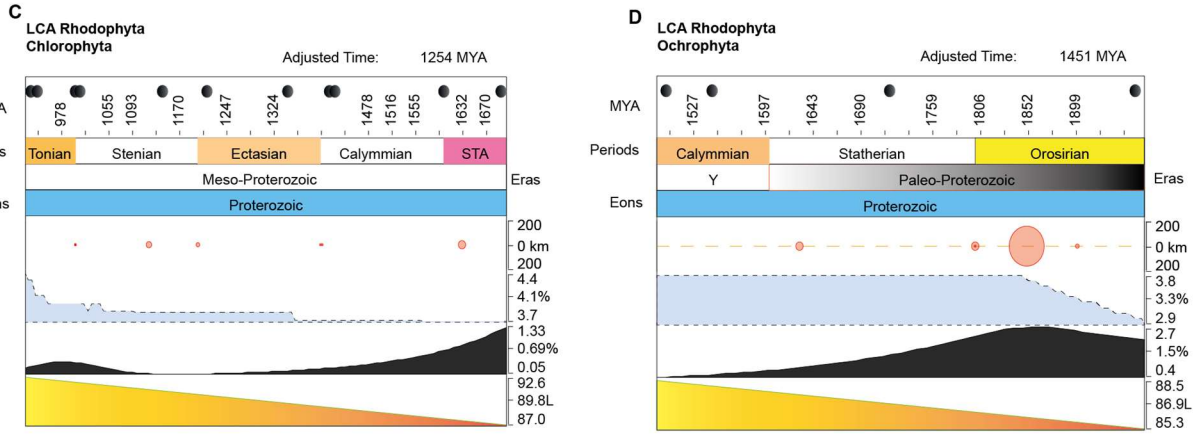
Supplemental information

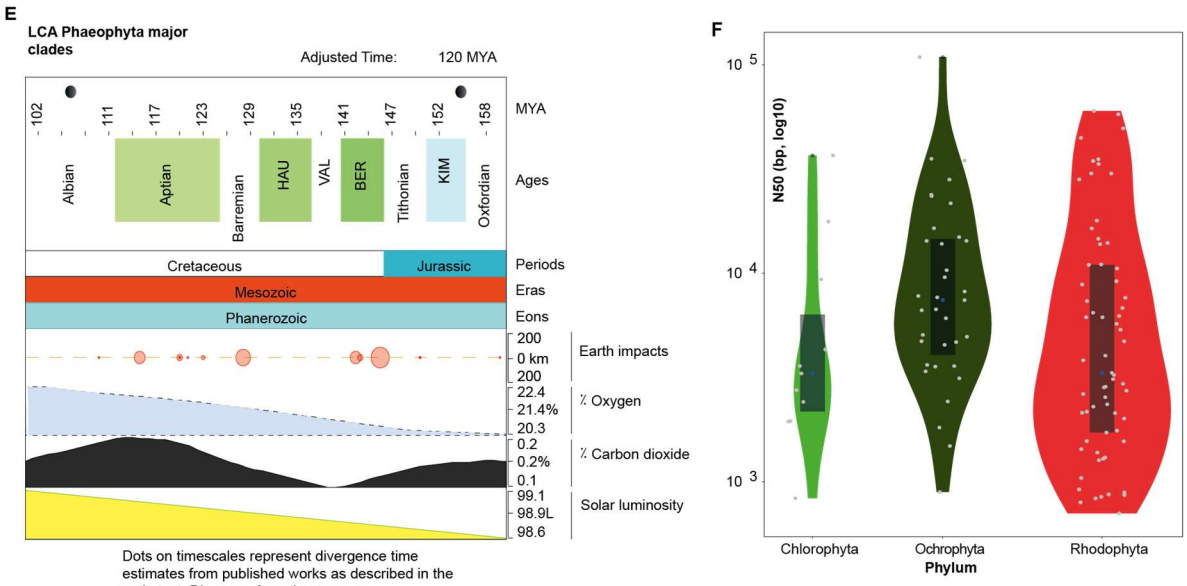
Macroalgal deep genomics illuminate multiple paths to aquatic, photosynthetic multicellularity

David R. Nelson, Alexandra Mystikou, Ashish Jaiswal, Cecilia Rad-Menendez, Michael J. Preston, Frederik De Boever, Diana C. El Assal, Sarah Daakour, Michael W. Lomas, Jean-Claude Twizere, David H. Green, William C. Ratcliff, and Kourosh Salehi-Ashtiani

SUPPLEMENTAL INFORMATION







Dots on timescales represent divergence time estimates from published works as described in the main text. Diagrams from timetree.org

Figure S1. Genome metrics and environmental and geological context for macroalgal evolution for the newly sequenced macroalgae. Related to Fig. 1.

(A) Phylogenetic tree from NCBI taxonomy showing the divergence of Rhodophyta, Chlorophyta, and, more recently, Ochrophyta lineages (Berney and Pawlowski, 2006; Blank, 2013; Caspermeyer, 2017; Douzery et al., 2004; Fiz-Palacios et al., 2011; Gaya et al., 2015; Gueidan et al., 2011; Heckman et al., 2001; Hedges et al., 2004; Herron et al., 2009; Lang et al., 2010; Munakata et al., 2016; Parfrey et al., 2011; Yang et al., 2016; Yoon et al., 2004; Zimmer et al., 2007). Collapsed tree branches indicate major taxa comprising our culture collection for genome sequencing (n = 110). The Phaeophyte seaweeds evolved much more recently than the other two clades.

(B) Flow diagram showing the morphological innovations characterizing multicellular algae from the three major lineages.

(C-E) Timelines for environmental and geological conditions during the evolution of the Chlorophyta, Rhodophyta, and Ochrophyta macroalgae. This figure was based on diagrams produced by timetree.org using referenced estimates of divergence from the LCAs of the algal phyla based on molecular clock experiments and geological records (Berney and Pawlowski, 2006; Blank, 2013; Caspermeyer, 2017; Douzery *et al.*, 2004; Fiz-Palacios *et al.*, 2011; Gaya *et al.*, 2015; Gueidan *et al.*, 2011; Heckman *et al.*, 2001; Hedges *et al.*, 2004; Herron *et al.*, 2009; Lang *et al.*, 2010; Munakata *et al.*, 2016; Parfrey *et al.*, 2011; Yang *et al.*, 2016; Yoon *et al.*, 2004; Zimmer *et al.*, 2007).

(F) N50 values for the initial macroalgal genome assemblies.

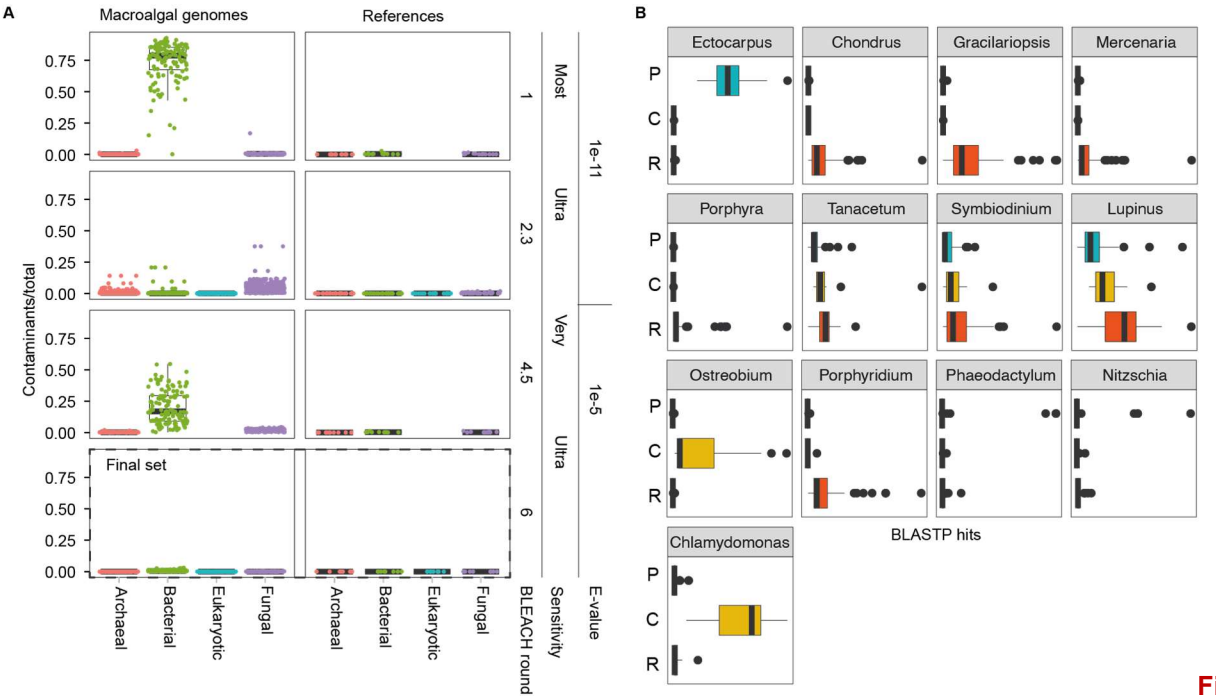
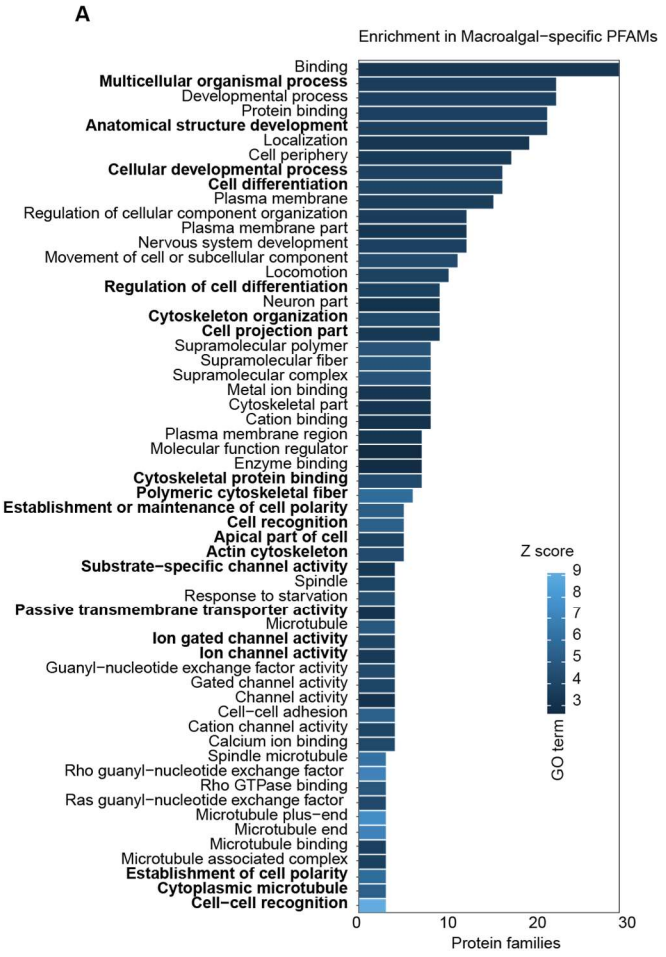
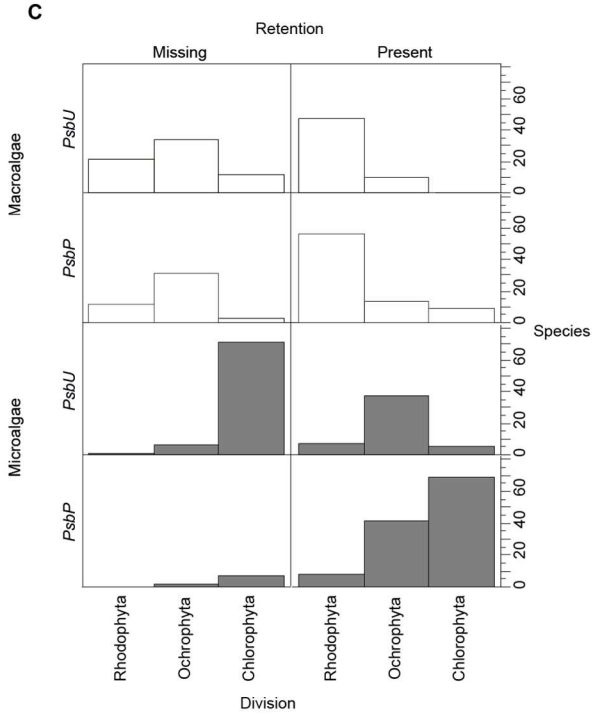


Figure S2. Decontamination statistics for the newly sequenced macroalgal genomes. This figure shows quantities of contaminant sequences removed in each of the BLEACH decontamination rounds. Each BLEACH round further reduced the effective genome size of the sequenced macroalgae, and eventually produced the working set of genomes roughly equivalently clean to the reference genomes. Related to Fig. 1.

(A) Ratio of contaminant sequences removed from each macroalgal assembly in iterative BLEACH runs. The assemblies were similar to reference genome assemblies in terms of contaminant sequences detected in later BLEACH runs.

(B) Classification of BLASTP hits from representative genomes into either Rhodophyta, Chlorophyta, or Ochrophyta.


A
Enrichment in Macroalgal−specific PFAMs
Binding
Multicellular organismal process
Developmental process
Protein binding
Anatomical structure development
Localization
Cell periphery
Cellular developmental process
Cell differentiation
Plasma membrane
Regulation of cellular component organization
Plasma membrane part
Nervous system development
Movement of cell or subcellular component
Locomotion
Regulation of cell differentiation
Neuron part
Cytoskeleton organization
Cell projection part
Supramolecular polymer
Supramolecular fiber
Supramolecular complex
Metal ion binding
Cytoskeletal part
Cation binding
Plasma membrane region
Molecular function regulator
Enzyme binding
Cytoskeletal protein binding
Polymeric cytoskeletal fiber
Establishment or maintenance of cell polarity
Cell recognition
Apical part of cell
Actin cytoskeleton
Substrate−specific channel activity
Spindle
Response to starvation
Passive transmembrane transporter activity
Microtubule
Ion gated channel activity
Ion channel activity
Guanyl−nucleotide exchange factor activity
Gated channel activity
Channel activity
Cell−cell adhesion
Cation channel activity
Calcium ion binding
Spindle microtubule
Rho guanyl−nucleotide exchange factor
Rho GTPase binding
Ras guanyl−nucleotide exchange factor
Microtubule plus−end
Microtubule end
Microtubule binding
Microtubule associated complex
Establishment of cell polarity
Cytoplasmic microtubule
Cell−cell recognition
Z score
9
8
7
6
5
4
3
GO term
0
10
20
30
Protein families


C
Retention
Missing
Present
Macroalgae
Microalgae
PsbU
PsbP
PsbU
PsbP
0
20
40
60
Species
Rhodophyta
Ochrophyta
Chlorophyta
Rhodophyta
Ochrophyta
Chlorophyta
Division

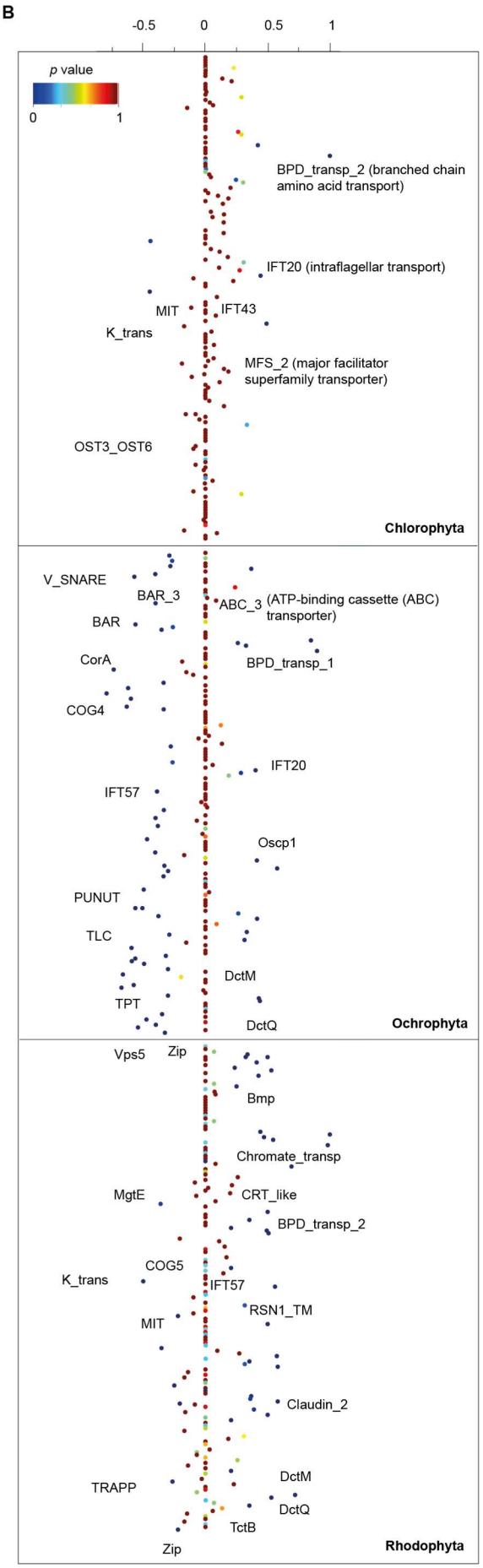

B
-0.5
0
0.5
1
p value
0
1
BPD_transp_2 (branched chain amino acid transport)
IFT20 (intraflagellar transport)
MIT
IFT43
K_trans
MFS_2 (major facilitator superfamily transporter)
OST3_OST6
Chlorophyta
V_SNARE
BAR_3
ABC_3 (ATP-binding cassette (ABC) transporter)
BAR
CorA
BPD_transp_1
COG4
IFT20
IFT57
Oscp1
PUNUT
TLC
DctM
TPT
DctQ
Ochrophyta
Vps5
Zip
Bmp
Chromate_transp
MgtE
CRT_like
BPD_transp_2
COG5
K_trans
IFT57
RSN1_TM
MIT
Claudin_2
TRAPP
DctM
DctQ
TctB
Zip
Rhodophyta

Figure S3. Macroalgal gain/loss of genes involved in photosynthesis and transport. Transport and transporter-related proteins were among the top gene families differentiating macroalgal genomes from those from their microalgal counterparts; photosynthetic component retention only sparsely varied (related to Fig. 3).

(A) Gene ontology (GO) enrichment in the set of 222 PFAMs unique to macroalgae compared to their microalgal counterparts.

(B) Transporter retention rates for micro-/macroalgae using microalgae as proxies for macroalgal unicellular ancestors. Negative values along the x-axis indicate PFAMs missing from macroalgae. Positive values reflect domains in macroalgae missing from microalgae. Most transporter PFAMs showed insignificant variation (values of zero on the x-axis).

(C) Photosystem II subunit PsbU and PsbP retention in the major macroalgal clades. The PsbU gene was generally conserved in microalgae and retained by only Rhodophyte macroalgae. These losses suggest its non-essentiality for multicellular Phaeophyte and Chlorophytes. Only Rhodophytes retained PsbU with substantial conservation.

A

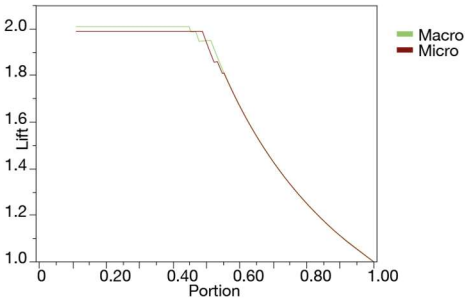
Measures	Value
Generalized RSquare	0.9515838
Entropy RSquare	0.9021709
RASE	0.140477
Mean Abs Dev	0.0474745
Misclassi fication Rate	0.0248756
-LogLikelihood	13.629554
Sum Freq	201

Confusion Matrix

Actual MicroMacro	Predicted Count Macro	Micro
Macro	98	2
Micro	3	98

Confusion Rates

Actual MicroMacro	Predicted Rate Macro	Micro
Macro	0.980	0.020
Micro	0.030	0.970



B

Measures	Value
Generalized RSquare	0.9287362
Entropy RSquare	0.8602661
RASE	0.1605528
Mean Abs Dev	0.057959
Misclassi fication Rate	0.0196078
-LogLikelihood	4.9382953
Sum Freq	51

Confusion Matrix

Actual MicroMacro	Predicted Count Macro	Micro
Macro	24	1
Micro	0	26

Confusion Rates

Actual MicroMacro	Predicted Rate Macro	Micro
Macro	0.960	0.040
Micro	0.000	1.000

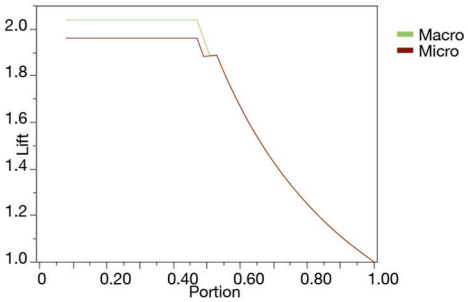


Figure S4. A PFAM-based aNN can distinguish between multi- and unicellular algae. Metrics for the 841TLGx24 aNN model's performance where (A) shows training metrics from the training set and (B) shows validation metrics from a holdout set of genomes. The full list of contributor PFAMs is in Table S4. Lift plots are at their far-right sides. These plots show the 'lift' provided by the variables used in the model. Briefly, lift values indicate the predictive power of different variable sample sizes. For example, an approximate lift of 2.0 is shown when only 40% of variables are used, roughly indicating the potency of the input variable set towards the model's predictive capacity. These metrics for the 841TLGx24 aNN model show performance on a set not used in training. The results indicate the model's performance with novel macroalgal genomic queries. We thus conclude that the sample size, cleanliness of data, and model parameters are sufficient to build an aNN model predictive of multicellularity status from the adhesome contents of new queries. Related to Fig. 4.

Figure S5. EsV-1-7 motif-containing genes and codomain architecture. (A-D) The EsV-1-7 domains were often seen in tandem repeats and in predicted membrane-localized proteins. Their ubiquitous, and recent, acquisition by Phaeophytes suggests their influence in the rapid expansion and dominance of many members from this phylum. Related to Fig. 6.

Table S1. Functional annotation and metadata for macroalgal species. This is a multi-sheet Excel workbook containing the PFAM count matrix for decontaminated assemblies and their strain metadata, contamination estimates, assembly metrics including BUSCO and N50 scores, and source data for the ternary analysis. Related to Figs. 2 and 3.

Table S2. GO enrichment in macroalgal-specific genes. This is a multi-sheet Excel workbook containing enriched GO terms in macroalgal-specific PFAMs conserved in Rhodophyta, Ochrophyta, and Chlorophyta and response screening results comparing means in PFAM counts between divisions. Related to Fig. 3.

Table S3. Comparative genomics of micro- and macroalgae. This is a multi-sheet Excel workbook containing response screening tables comparing PFAM and GO variation in micro- compared to macroalgae. Related to Fig. 4.

Table S4. Unraveling the macroalgal adhesome. This is a multi-sheet Excel workbook containing macroalgal adhesome atlas and response screens among macroalgal phyla and between micro- and macroalgae. Related to Fig. 4.

Table S5. Endogenous viral elements in macroalgae. This is a multi-sheet Excel workbook containing VFAM count matrix and response screening results for comparisons of macroalgal VFAM counts by climate and habitat. This table also includes the EVOP matrix and response screening results comparing EVOPs found in the macroalgal genomes among climates and macroalgal tORFs with EsV-1-7 domains and their codomains. Related to Fig. 5.

Data S1. Original and decontaminated macroalgal assemblies from macroalgae cultured from international culture collection centers. These data are available in compressed tar archives at Zenodo.org (doi:10.5281/zenodo.7758534, https://zenodo.org/record/7758534). Assemblies used in this study from published work and from local macroalgae isolates are not included in this dataset. Related to all figures.

Data S2. Macroalgae genome annotations. The original (~52 GB) and decontaminated (~7 GB) gene annotations, including coding sequences (CDSs), proteins (AA), and coordinate (GFF) files (doi:10.5281/zenodo.7751045, https://zenodo.org/record/7751045). Related to all figures.

Data S3. Analysis data files. This dataset includes data files for the analyses presented in the manuscript, including:

(A) Decontamination analysis, including iterative BLEACH contamination calls, GFF coordinates for contaminants, and downsampling analyses of decontaminated genomes. Related to Fig. S1.
(B) HMMsearch results for decontaminated assemblies (PFAMs). Related to Figs. 2-6.
(C) Ternary analyses including dcGO enrichment for >80% purity sets for the three phyla. Related to Fig. 2.
(D) Comparative genomics analyses of micro- and macroalgal genomes, including intersection, response screening, metabolic pathway, GO enrichment, and aNN modeling analyses. Related to Fig. 3.
(E) Adhesome analysis including HMMsearch results for adhesome domains and codomains and response screening analyses between phyla, habitat, climate, and micro- vs. macroalgae. The neural network model using the 110 adhesome PFAMs is also included in this dataset. Related to Fig. 4.
(F) Endogenous viral element analyses, including VFAM HMMsearch results, EVOPs, macroalgal sequences with EsV-1-7 codomains and comparative analyses including response screens and hierarchical clustering results. Related to Fig. 5.
(G) All computational scripts used for analyses in this study. Scripts are either '.sh' or '.sbatch' files for execution in a Linux environment with a SLURM (https://github.com/SchedMD/slurm) high-performance computing (HPC) scheduler. Related to all analyses.

These data are available in compressed tar archives at Zenodo.org (doi:10.5281/zenodo.7947301, https://zenodo.org/record/7947301).

Data S4. Supplementary Tables. This dataset includes all supplemental tables (Tables S1-S5, doi: 10.5281/zenodo.10581453, https://zenodo.org/records/10581453).

Note S1 - Expanded technical and contextual descriptions of analyses

(A) PFAMs and GO enrichment differentiating micro- from macroalgae

These PFAMs were found in members of each of the three macroalgal phyla but not in any members of their corresponding microalgal groups. These 222 macroalgal-specific PFAMs were used in a domain-based analysis to recover 59 GO terms. These 59 GO terms essentially provide a genetic basis among aquatic multicellular algae differentiating them from their unicellular relatives. Prominent PFAMs in this group were family 6 glycosyltransferases, retinol binding protein receptors, cell-cycle alteration and expression-elevated proteins, cation-independent mannose-6-phosphate receptors, growth-arrest-specific Protein 2, acid sphingomyelin phosphodiesterase, Glyoxalase 8, and a

COMMD1 protein were found in macroalgal genomes from all three phyla, but not in the microalgal genomes from the corresponding three phyla.

The top GO term enriched for in the *n* = 222 group was 'Rho guanyl-nucleotide exchange factor activity' (Z = 7.02, *adj. p* = 1.47 x 10^{-2}). Rho GTPase controls actin dynamics and is essential for cell differentiation and differential adhesion networks.(Nayak et al., 2013) This GO term was comprised of the PFAMs 'Regulator of G protein signaling-like domain (PF09128) and DOCK_N (PF16172), which is an N-terminal α-helix domain of the Dedicator Of CytoKinesis (DOCK) protein. DOCK proteins are guanine nucleotide exchange factors (GEFs) for RHO GTPases and influence intracellular actin dynamics, such as those utilized by various linker proteins in complex tissue formation and organization (Benson and Southgate, 2021; Hernandez-Vasquez et al., 2017; Willoughby et al., 2017). The DOCK_N domain is situated between a C2 domain and a variable SH3 domain and facilitates interactions between ELMO subunits and SH3 and ARM domains to destabilize auto-inhibition and activate DOCK (Chang et al., 2020). These proteins may play important roles at the macroalgal cell membrane to coordinate communication among different cell types.

The GO terms 'regulation of cell projection' and 'cytoskeleton organization' were enriched in the macroalgal-specific genes. Cellular reorganization in differentiated cells is essential for a multicellular phenotype. 'Cell-cell recognition' (GO:0009988, Z = 9.11, *adj. p* = 2.54 x 10^{-2}) includes the protein family PF00769 (Ezrin/radixin/moesin family C terminal – ERM_C), which links actin filaments of cell surface structures to the plasma membrane (Yonemura et al., 1998). The extracellular matrix (ECM) and filopodia, thin, actin-rich plasma-membrane protrusions, promote the cohesion of animal blastulae (Brunet and King, 2017) and are necessary for embryonic development (Fierro-Gonzalez et al., 2013). Other actin-regulated proteins, such as Profilin-1 (Pfn1), seen in our data, are vital in cell type specialization; remodeling cytoskeletons to extreme ends, such as in osteoclasts, and modulating cellular migration patterns (Kajikawa et al., 2022). We saw significant increases in Spondin domains in macro- compared to microalgae (adj. p in Rhodophytes = 1.35 x 10^{-174}, Chlorophytes = 6.64 x 10^{-163}). Still, only 85/128 macroalgae surveyed had at least one Spondin copy. Ochrophyte macroalgae did not have significantly more spondins than their unicellular correlates. The reference Rhodophyte genomes were among the top Spondin-rich genomes (e.g., *Gracilariopsis chorda*, 15 domains; *Chondrus crispus*, 12 domains). Spondins regulate cell-cell adhesion and outgrowth; their cleavage and subsequent membrane docking generate combinatorial guidance cues for cell outgrowths (Burstyn-Cohen et al., 1999; Klar et al., 1992; Zisman et al., 2007). These morphology-guiding genes appear to be keystone elements in the Chlorophyta and Rhodophyta macroalgal genomes, differentiating them from their microalgal correlates.

(B) Expanded technical and contextual descriptions of macroalgal ECM analyses

The development of intricate ECMs, which can play an important role in multicellularity (Brunet and King, 2017), enabled cell connection, positioning, and defense in macroalgae (Kloareg et al., 2021). These ECMs facilitated the creation of complex, adaptive structures (e.g., thallus, holdfast rhizoid, large blades, and bladders), providing mechanotransduction for tide-resilient tissues (Humphrey et al., 2014) and enabling small molecule transport for nutrient supply. While simple multicellular lineages share these processes, macroalgae exhibit direct cell-cell contacts via lineage-specific channels like plasmodesmata (Nagasato et al., 2022; Terauchi et al., 2015) thought to have evolved independently across macroalgal lineages. Membrane topology-altering proteins are crucial for plasmodesmata formation, with diverse proteins forming distinct membrane morphologies. The Phaeophyta/Chlorophyta specific sets of ankyrins (Fig. 1A) may play important roles in plasmodesmata formation in these lineages. Macroalgal-specific PFAMs, conserved in all three phyla, were enriched for the GO term 'regulation of cytoskeleton organization' (GO:0008092, Z = 4.22, adj. p = 1.62 x 10^{-2}), suggesting conservation of genes guiding morphology through cytoskeletal rearrangements.

Polysaccharides and proteoglycans, attached to a wide variety of glycoproteins, are abundant in the ECM and are essential for tissue generation and maintenance. The macroalgal-specific gene set conserved in all three lineages (n = 222) included a significant representation of genes in polysaccharide and proteoglycan metabolic pathways with putative roles in ECM development (Fig. 3). For example, the lipopolysaccharide biosynthesis reaction catalyzed by a UDP-N-acetylglucosamine acyltransferase (EC: 2.3.1.129) was uniquely present in macroalgae. This enzyme works with a lipid-A-disaccharide synthase (EC: 2.4.1.182) and a tetraacyldisaccharide kinase (EC: 2.7.1.130) to synthesize phosphorylated glycolipids (Anderson et al., 1985). Mucin-type O-glycan biosynthesis, catalyzed by an N-acetylgalactosaminide beta-1,3-galactosyltransferase (NABG, EC: 2.4.1.122), was also represented by enzymes unique to macroalgae. The NABG protein acts on non-reducing O-serine-linked N-acetylgalactosamine residues in mucin glycoproteins, forming antigen proteins (Hesford et al., 1981). These molecules are essential in the human immune system, relaying signals of neighboring cell detection. Although their roles in maintaining a multicellular phenotype in macroalgae have not yet been explored, their unique retention in these lineages compared to their microalgal counterparts hints at a antigenicity cell-cell recognition system in the macroalgae.

(C) Expanded technical and contextual descriptions of macroalgal adhesome accessory protein families

Adhesome accessory, or codomains, were mostly distributed with sparse evolutionary patterns. In macroalgal cadherins, gene architecture was highly variable, with juxtamembrane regions (JXRs) placed at variable lengths from predicted membrane-bound regions of the tORFs analyzed. The JXR determines the strength of cadherin binding (Evdokimov et al., 2016; Paulson et al., 2000; Yap et al., 1998), so its variable placement would be expected to influence the specificity of differential tissue interface. The JXR is the cytoplasmic region adjacent to the cadherin-occupied transmembrane space (COTS) (Evdokimov *et al.*, 2016; Paulson *et al.*, 2000; Yap *et al.*, 1998) and regulates cell-cell binding (i.e. adhesion) in the extracellular domain of E-cadherin through cytoskeletal interactions (Nagafuchi and Takeichi, 1988). Its variable presence in macroalgal cadherins implies a wide variety of cell-cell binding modes. The activity of this domain also leads to cell clustering (Yap *et al.*, 1998) which is a prerequisite for multicellularity (Ratcliff et al., 2012). Rhodophyte macroalgae had more E-cadherins compared to their microalgal counterparts (PF00028, CL0159, adj. $p = 0.000063$). These glycoproteins extend through the cell membrane to connect with cadherins on other cells (adhesive *trans* binding) or laterally in signal-transducing *cis* binding with nearby cadherins on the same cell. Recent superresolution microscopy and intermolecular single-molecule Förster resonance energy transfer experiments showed mutual cooperativity of cis and trans cadherin interactions towards cell-cell adhesion (Thompson et al., 2021), hinting at an unexplored meta layer of cell-cell interactions. Adhesive networks, generated by combinations of cadherin cis/trans interactions in determining density patterns at the cell exterior interface, serve as morphological blueprints for tissue organization and overall organismal body plans. The recovered collection of macroalgal cadherins and codomains outlines this facet of their epochal-dependent approach to multicellular development.

Various lectins were cadherin codomains. Lectins regulate glycoprotein (e.g., ECM components) and cell adhesion, among other processes (Cooper and Barondes, 1990; Derouiche et al., 1996; Ko, 1987; Moulessehoul et al., 1992; Neu and Kuhlicke, 2017; Outenreath et al., 1988). Plant lectins can recognize various hydrophobic ligands, including indole acetic acid (IAA), cytokinin, adenine, auxins, and water-soluble porphyrins (De Coninck and Van Damme, 2021; Ma et al., 2021; Ricci-Azevedo et al., 2017). Lectin C domains (PF00059, clan CL0056) were significantly increased in macroalgal compared to microalgal Rhodophytes (adj. $p = 0.031$); the expansion of these genes may constitute a unique facet of cell-cell adhesion in the Rhodophyta. Lectin_C requires calcium for activation. Calcium signaling is believed to function in multicellular algae similar to plants and metazoans, with the Characean macroalga *Nitellopsis obtusa* serving as a basal model for action potential excitation of Ca^{2+} and K^{+} channels (Koselski et al., 2021). Lectin_C N-acetylglucosamine receptors facilitate receptor-mediated endocytosis (Falasca et al., 1992; Singh et al., 2014). N-Acetylglucosamine and related sugars are responsible for the cell-to-cell communication mechanisms vital to multicellular organisms. For example, endogenous GlcNAcs control multicellular filamentous growth in response to temperature as a highly conserved cue of morphogenesis in fungi (Gilmore et al., 2013). In *Candida albicans*, GlcNAcs mediate the shift from a parasitic to environmental lifestyle. The acquisition of sophisticated channel-coupled receptor networks appears to have demarcated the rise into multicellularity for the Chlorophyta macroalgae. Other transmembrane cadherin codomains were membrane-bound extracellular proteins like the sulfur-rich EGF (e.g., EGF, Laminin_EGF/G1/2/3, domains, which include six cysteine residues constituting three disulphide bonds). Cadherin-EGF proteins may mediate signaling and adhesion with sulfated polysaccharides (e.g., alginates) in macroalgal ECMs.

Transmembrane ion channels were integrin codomains in macroalgal genes. Known integrin-channel complexes regulate excitation-contraction coupling, non-associative learning, and embryogenesis (Forzisi and Sesti, 2022). We saw complement control protein (CCP) modules that were concurrent to cadherin domains in ORFs. For example, SUSHI domains were found in the same CDS as macroalgal cadherins. The SUSHI domain has a conserved 60 amino acid track with four cysteine residues necessary for disulfide-bridges essential for its folding (Libert and Vassart, 1989; Reid and Day, 1989). SUSHI proteins are essential for various cell-cell recognition processes (Wei et al., 2001). For example, the tobacco hornworm *Manduca sexta* pattern recognition serine proteinase, which triggers a prophenoloxidase activation cascade, contains a Sushi domain and a unique Cys-rich region in the same CDS (Ji et al., 2004). Other cadherin codomains in the macroalgae included 7tm_2 domains which are transmembrane domains in the secretin-receptor family of the G-protein-coupled receptors. Cadherin-7tm_2 codomains indicate cell-cell signaling through GPCRs proteins for cells in direct contact. Lectin domains, including Lectin_C, were cadherin codomains. Malectins were found adjacent to cadherin domains in several macroalgal genomes. Malectins are membrane-anchored proteins of the endoplasmic reticulum that recognize and binds Glc2-N-glycans (Schallus et al., 2008).

We showed that a distinct set of polysaccharide metabolic pathways differentiated macroalgae from microalgae (Fig. 3D-F). Cadherin codomains also included a wide variety of polysaccharide metabolic genes, including sugar isomerases and epimerases necessary for ECM construction. The beta-propeller fold glucose / sorbosone dehydrogenase Kringle domain was found on a macroalgal cadherin. Kringles are disulfide-, beta sheet-rich proteins that promote blood clotting and regulate fibrinolytic proteins in metazoans (Ozhogina et al., 2001). They are in plasminogen, hepatocyte growth factors, prothrombin, and apolipoprotein A, and have been shown to bind to integrins (Takada, 2012). Their presence in macroalgae suggests their adoption as a co-binding domain to supplement the role of this cadherin as a cellular aggregation factor. Integrins are also ubiquitous in multicellular organisms, and evidence suggest their activities and diversity influence organismal complexity and resultant morphology by way of tissue-specific adhesion mechanisms

(Green and Brown, 2019; Song et al., 2006; Tang et al., 2008; Tong et al., 2018; Zarbock et al., 2007). Other integrin cadherin codomains were uncovered. For example, FG-GAP domains, the extracellular alpha integrin repeats, were found on cadherins. In conjunction with the other differential binding domains, their presence suggests a mix-and-match approach to intercellular binding in macroalgal tissues. Exotic combinations of glycoprotein linkages with adaptable tensile strengths would support tissue integrity in various oceanic conditions presenting high shear forces to their inhabitants.

We observed calx-beta motifs concurrent with cadherins. Calx-beta PFAM domains mediate the bi-directional signal transfer across the plasma membrane from the cytoplasmic tail of integrins (Schwarz and Benzer, 1997). Their presence as codomains in macroalgal cadherins suggests yet another route to modulating cell-cell adhesion or coordinated aggregation. Mat-forming marine cyanobacteria use Calx-beta domains in aggregative multicellularity (Schwarz and Benzer, 1999). Calx domains are necessary for sodium-calcium exchangers that export calcium. Peripheral tissues in macroalgae have high sodium but low calcium exposure, and vice versa for inner tissues. These domains are more likely to indicate a linkage between inner and outer macroalgal tissues. Similar systemic ion balancing mechanisms function in plants, facilitating ionic harmony between the static substratum (e.g., soil, rock, or coral) and flowing stratum (e.g., seawater or periodic rains and humidity).

Integrins dynamically regulate cell-cell adhesion through interactions with other integrins and various ECM components and other proteins to form integrin adhesion complexes (IAPs) (Campbell and Humphries, 2011; Zheng and Leftheris, 2020). Integrin and cadherin networks cooperate to form functional ECMs, and the integrin adhesome is equivalently complex to the cadhesome (Winograd-Katz et al., 2014; Zaidel-Bar and Geiger, 2010; Zaidel-Bar et al., 2007). The IAPs transduce signals that provide information on cell location, strata, and neighboring cell types. Variation in structure gives rise to two forms, arch and hand-gun, that can move through fly-casting, switchblade, and deadbolt mechanisms (Campbell and Humphries, 2011). Genetic ablation of integrin expression blocks embryogenesis in multicellular organisms (Green and Brown, 2019). The gain of integrin genes with various codomains in macroalgal progenitor lineages highlights their vital incorporation at the sunrise of macroalgal multicellular evolution. Each of the three major macroalgal lineages we examined had distinct integrin collections. Each integrin can bind multiple kinds of ECM components, and ECM components can bind multiple integrins (Campbell and Humphries, 2011). Therefore, the combinatorial complexity made possible by dynamic integrin-ECMs lends these interactions to a wide variety of cellular, tissue-specific, and organismal functions.

(D) Expanded technical and contextual descriptions of global VFAM analyses

Our results suggest that broadly conserved viral element sequence integrations denote ancient transfers, while sparse, stochastic viral element accumulation in a few genomes per clade indicates more recent and transitory events (Fig. 5). We observed drastic differences in VFAM content in Rhodophyta from different habitats (Fig. 5G), with little to no variance in Chlorophyta or Phaeophyta. These VFAMs included VFAM_964 from *Ictalurivirus*, or fish virus, family, and VFAM_883, an *Iridiovirus*, or insect virus, HMM. Rhodophyta macroalgae have been prey for multitudes of fish and aquatic insect species for more than a billion years of evolution; thus, these habitat dependent VFAMs may originate from viruses of macroalgal predators. Evidence for cross-kingdom viral infection is scant, but our data provide further evidence of the potentially broad host range of ancient marine viruses. Very-broad host range viruses could have exploded across the globe in rapidly-spiking pandemics affecting multiple species. Still, their genetic material would not have survived unless they provided some kind of benefit to their host lineage. This manner of horizontal spreading of genetic material at defined, relatively short geological intervals may be the source of the rampant viral sequence insertions we see.

Genes coding for viral-origin proteins were found in extensive tandem duplication arrays in macroalgal genomes, indicating large-scale genetic invasion by, or cooption of, viral DNA fragments as well as duplication of these elements. The Phaeophyte *Ectocarpus siliculosus* has many, characterized genomic viral sequence insertions (Cock et al., 2010a; Cock et al., 2010b; Cormier et al., 2017; Dittami et al., 2020); here, we show that viral elements are common to Rhodophyta and Chlorophyta in addition to other members of the Ochrophyta. Common viral genes with an apparent tendency to incorporate into their macroalgal hosts' genomes were those for potassium channels, which have also been inserted into Chlorophyta genomes (Chen et al., 2005), and repeated copies of ankyrin genes (Mahmood et al., 2012), which have been hypothesized to increase algal survival in seawater by reinforcing cell membrane architecture.(Nelson et al., 2021) Virus-type protein families were highly represented in total PFAM counts in members from all lineages; however, reverse transcriptase (RT)-like domains were remarkably abundant in Rhodophyta and Chlorophyta genomes compared to those of Pheaophyta. These domains may perform core genome maintenance functions (Brown and Freudenreich, 2021; Chan and Wong, 2012; Hartley and O'Neill, 2019).

Our VFAM analyses showed that clade-wide retention of adaptive virus-origin genes partly defines species and ecotypes. For example, *Gracilaria* species preferentially occupy nutrient-rich, cool temperate waters. Our dataset reveals that their VFAM profile is distinctive to their genera. Although the integration timelines in many macroalgal later-branching clades are relatively recent, several VFAM explosions in the Phaeophyte algae are as current as within the last few ice ages (Table

S5). Interestingly, polar isolates had significantly higher levels of VFAMs in their genomes than subpolar isolates, hinting at a possible VFAM endogenization response to glaciation cycles. The Phaeophyta are liable to be infected with lysogenic viruses, unlike Rhodophyta and Chlorophyta, whose members have high infection rates by lytic viruses (Horas et al., 2018). Recent findings suggest that genes enabling multicellularity in Phaeophyceae may have evolved because of viral interactions. Certain Phaeophyte motifs, including the EsV-1-7 cysteine-rich domain Esv-1-7 (PF119, are not seen in other Stramenopiles or macroalgae phyla but are commonly seen in *Ectocarpus siliculosus virus* 1 (EsV-1) genomes (Delaroque et al., 2001). Phaeophyte developmental phenotypes can be drastically altered by modification of EsV-1-7 motif-rich genes. For example, the Immediate Upright (IMM) gene product of *Ectocarpus* (Peters et al., 2008) contains a 38 amino acid sequence similar to a cysteine-rich repeat motif from the EsV-1-7 gene in one of its main viruses, EsV-1(Delaroque *et al.*, 2001). Mutation of IMM resulted in reduced rhizoid growth and increased upright sporophyte filament growth, drastically altering the *Ectocarpus* developmental program. The virus has five unique cysteine-rich domains (Macaisne et al., 2017), but many more copies were interspersed with algal genes in the macroalgal genomes. Long tracks of up to 20 EsV-1-7 motifs per coding sequence were seen in tORFs with distinct, short transmembrane regions (Fig S5 - EsV-1-7 gene motifs). These common structures in EsV-1-7 motif-containing genes hint at their functionality in cell-cell interactions.

(E) Expanded descriptions of EsV-1-7 codomains and gene structures

Catalytic domains for RNA modification were seen in EsV-1-7 CDSs. A QueF domain (PF14489.9, E = 8.5 x 10^{-6}), initially described in Lanen *et al.*)(Van Lanen et al., 2005) was found in EsV-1-7 CDSs in our dataset (see Table S5 and Data S3). This domain facilitates the biosynthesis of queuosine, which is a structurally complex, non-canonical RNA nucleoside serving as a modification in tRNATyr, tRNAHis, tRNAAsn, and tRNAAsp (Hillmeier et al., 2021; Zhang and Pan, 2021). The acquisition of QueF domains (e.g., from a virus to its host alga) could introduce novel levels of complexity in coding potential during multicellular development, such as that seen in *Dictyostelium discoideum* and many other eukaryotes (Hoffmann et al., 2021). Interestingly, keto-disaccharide hydrolase (keto-disac) domains were found in Phaeophyceae EsV-1-7 genes. This PFAM catalyzes the isomerization of disaccharides, including ellobiose, palatinose, maltose, sucrose, isomaltose, melibiose, and trehalose (Gao et al., 2016). These unique disaccharides, as potential ECM building blocks, are subject to environmental selection. The presence of a catalytic domain for sugar isomerization alongside cadherin domains implies previously undiscovered cell-to-cell linkage networks may depend on such multifunctional proteins. Cadherins and related cell-cell linker proteins, often glycoproteins with modifiable sugar arrays, suggest that the viral-origin keto-disaccharide hydrolases in macroalgal genomes could have provided host algae with novel cell-cell connection mechanisms.

Various ankyrin domains were observed in EsV-1-7-containing ORFs. For example, Ank (PF00023.33, 5.8 x 10^{-8}), Ank_2 (PF12796.10, 3.4 x 10^{-10}), and Ank_4 (PF13637.9, E = 4.9 x 10^{-7}) genes had EsV-1-7 domains. Evidence for the insertion of multiple copies of ankyrin genes from viruses into their hosts' genomes has been observed in microalgae (Nelson *et al.*, 2021) and Chlorophyta macroalgae (Fig. 1A). The saposin domain SapA (PF02199.18, 3 x 10^{-7}) was adjacent to some macroalgal EsV-1-7 domains. Saposins isolate lipid substrates for lysosomal degradation and exist in various forms in organisms spanning several kingdoms. They act to rearrange lipids, a process essential for cell specialization in multicellular organisms (Atilla-Gokcumen et al., 2014). Interestingly, plants have a plant-specific insert in their saposins which is thought to boost pathogen defense (Cheung et al., 2020). Lipid rafts, plasma membrane microdomains enriched in species-specific glycosphingolipids, gangliosides, and sterols, interact with cytoskeleton components to regulate morphology and signaling (Head et al., 2014). Thus, a viral-sourced saposin providing a means to produce daughter cells with altered morphology or signaling could yield an environment-dependent fitness boost from adopting a multicellular phenotype.

SUPPLEMENTAL INFORMATION REFERENCES

Anderson, M.S., Bulawa, C.E., and Raetz, C.R. (1985). The biosynthesis of gram-negative endotoxin. Formation of lipid A precursors from UDP-GlcNAc in extracts of Escherichia coli. The Journal of biological chemistry **260**:15536-15541.

Atilla-Gokcumen, G.E., Muro, E., Relat-Goberna, J., Sasse, S., Bedigian, A., Coughlin, M.L., Garcia-Manyes, S., and Eggert, U.S. (2014). Dividing cells regulate their lipid composition and localization. Cell **156**:428-439. 10.1016/j.cell.2013.12.015.

Benson, C.E., and Southgate, L. (2021). The DOCK protein family in vascular development and disease. Angiogenesis **24**:417-433. 10.1007/s10456-021-09768-8.

Berney, C., and Pawlowski, J. (2006). A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. Proc Biol Sci **273**:1867-1872. 10.1098/rspb.2006.3537.

Blank, C.E. (2013). Origin and early evolution of photosynthetic eukaryotes in freshwater environments: reinterpreting proterozoic paleobiology and biogeochemical processes in light of trait evolution. J Phycol **49**:1040-1055. 10.1111/jpy.12111.

Brown, R.E., and Freudenreich, C.H. (2021). Structure-forming repeats and their impact on genome stability. Curr Opin Genet Dev **67**:41-51. 10.1016/j.gde.2020.10.006.

Brunet, T., and King, N. (2017). The Origin of Animal Multicellularity and Cell Differentiation. Dev Cell **43**:124-140. 10.1016/j.devcel.2017.09.016.

Burstyn-Cohen, T., Tzarfaty, V., Frumkin, A., Feinstein, Y., Stoeckli, E., and Klar, A. (1999). F-Spondin is required for accurate pathfinding of commissural axons at the floor plate. Neuron **23**:233-246. 10.1016/s0896-6273(00)80776-x.

Campbell, I.D., and Humphries, M.J. (2011). Integrin structure, activation, and interactions. Cold Spring Harb Perspect Biol **3**10.1101/cshperspect.a004994.

Caspermeyer, J. (2017). Next Generation TimeTree: An Expanded History of Life on Earth at Your Fingertips. Mol Biol Evol **34**:1822-1823. 10.1093/molbev/msx133.

Chan, F.L., and Wong, L.H. (2012). Transcription in the maintenance of centromere chromatin identity. Nucleic Acids Res **40**:11178-11188. 10.1093/nar/gks921.

Chang, L., Yang, J., Jo, C.H., Boland, A., Zhang, Z., McLaughlin, S.H., Abu-Thuraia, A., Killoran, R.C., Smith, M.J., Cote, J.F., et al. (2020). Structure of the DOCK2-ELMO1 complex provides insights into regulation of the auto-inhibited state. Nat Commun **11**:3464. 10.1038/s41467-020-17271-9.

Chen, J., Cassar, S.C., Zhang, D., and Gopalakrishnan, M. (2005). A novel potassium channel encoded by Ectocarpus siliculosus virus. Biochem Biophys Res Commun **326**:887-893. 10.1016/j.bbrc.2004.11.125.

Cheung, L.K.Y., Dupuis, J.H., Dee, D.R., Bryksa, B.C., and Yada, R.Y. (2020). Roles of Plant-Specific Inserts in Plant Defense. Trends Plant Sci **25**:682-694. 10.1016/j.tplants.2020.02.009.

Cock, J.M., Coelho, S.M., Brownlee, C., and Taylor, A.R. (2010a). The Ectocarpus genome sequence: insights into brown algal biology and the evolutionary diversity of the eukaryotes. New Phytol **188**:1-4. 10.1111/j.1469-8137.2010.03454.x.

Cock, J.M., Sterck, L., Rouze, P., Scornet, D., Allen, A.E., Amoutzias, G., Anthouard, V., Artiguenave, F., Aury, J.M., Badger, J.H., et al. (2010b). The Ectocarpus genome and the independent evolution of multicellularity in brown algae. Nature **465**:617-621. 10.1038/nature09016.

Cooper, D.N., and Barondes, S.H. (1990). Evidence for export of a muscle lectin from cytosol to extracellular matrix and for a novel secretory mechanism. J Cell Biol **110**:1681-1691. 10.1083/jcb.110.5.1681.

Cormier, A., Avia, K., Sterck, L., Derrien, T., Wucher, V., Andres, G., Monsoor, M., Godfroy, O., Lipinska, A., Perrineau, M.M., et al. (2017). Re-annotation, improved large-scale assembly and establishment of a catalogue of noncoding loci for the genome of the model brown alga Ectocarpus. New Phytol **214**:219-232. 10.1111/nph.14321.

De Coninck, T., and Van Damme, E.J.M. (2021). Review: The multiple roles of plant lectins. Plant Sci **313**:111096. 10.1016/j.plantsci.2021.111096.

Delaroque, N., Muller, D.G., Bothe, G., Pohl, T., Knippers, R., and Boland, W. (2001). The complete DNA sequence of the Ectocarpus siliculosus Virus EsV-1 genome. Virology **287**:112-132. 10.1006/viro.2001.1028.

Derouiche, A., Hartig, W., Brauer, K., and Bruckner, G. (1996). Spatial relationship of lectin-labelled extracellular matrix and glutamine synthetase-immunoreactive astrocytes in rat cortical forebrain regions. J Anat **189 (Pt 2)**:363-372.

Dittami, S.M., Corre, E., Brillet-Gueguen, L., Lipinska, A.P., Pontoizeau, N., Aite, M., Avia, K., Caron, C., Cho, C.H., Collen, J., et al. (2020). The genome of Ectocarpus subulatus - A highly stress-tolerant brown alga. Mar Genomics **52**:100740. 10.1016/j.margen.2020.100740.

Douzery, E.J., Snell, E.A., Bapteste, E., Delsuc, F., and Philippe, H. (2004). The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? Proc Natl Acad Sci U S A **101**:15386-15391. 10.1073/pnas.0403984101.

Evdokimov, K., Biswas, S., Schledzewski, K., Winkler, M., Gorzelanny, C., Schneider, S.W., Goerdt, S., and Geraud, C. (2016). Leda-1/Pianp is targeted to the basolateral plasma membrane by a distinct intracellular juxtamembrane region and modulates barrier properties and E-Cadherin processing. Biochem Biophys Res Commun **475**:342-349. 10.1016/j.bbrc.2016.05.092.

Falasca, L., Lentini, A., and Dini, L. (1992). Receptor mediated endocytosis of N-acetylglucosamine and mannose exposing molecules by cultured chick embryo hepatocytes. Cell Mol Biol (Noisy-le-grand) **38**:621-627.

Fierro-Gonzalez, J.C., White, M.D., Silva, J.C., and Plachta, N. (2013). Cadherin-dependent filopodia control preimplantation embryo compaction. Nat Cell Biol **15**:1424-1433. 10.1038/ncb2875.

Fiz-Palacios, O., Schneider, H., Heinrichs, J., and Savolainen, V. (2011). Diversification of land plants: insights from a family-level phylogenetic analysis. BMC Evol Biol **11**:341. 10.1186/1471-2148-11-341.

Forzisi, E., and Sesti, F. (2022). Non-conducting functions of ion channels: The case of integrin-ion channel complexes. Channels (Austin) **16**:185-197. 10.1080/19336950.2022.2108565.

Gao, D.M., Kobayashi, T., and Adachi, S. (2016). Production of keto-disaccharides from aldo-disaccharides in subcritical aqueous ethanol. Biosci Biotechnol Biochem **80**:998-1005. 10.1080/09168451.2015.1127135.

Gaya, E., Fernandez-Brime, S., Vargas, R., Lachlan, R.F., Gueidan, C., Ramirez-Mejia, M., and Lutzoni, F. (2015). The adaptive radiation of lichen-forming Teloschistaceae is associated with sunscreening pigments and a bark-to-rock substrate shift. Proc Natl Acad Sci U S A **112**:11600-11605. 10.1073/pnas.1507072112.

Gilmore, S.A., Naseem, S., Konopka, J.B., and Sil, A. (2013). N-acetylglucosamine (GlcNAc) triggers a rapid, temperature-responsive morphogenetic program in thermally dimorphic fungi. PLoS Genet **9**:e1003799. 10.1371/journal.pgen.1003799.

Green, H.J., and Brown, N.H. (2019). Integrin intracellular machinery in action. Exp Cell Res **378**:226-231. 10.1016/j.yexcr.2019.03.011.

Gueidan, C., Ruibal, C., de Hoog, G.S., and Schneider, H. (2011). Rock-inhabiting fungi originated during periods of dry climate in the late Devonian and middle Triassic. Fungal Biol **115**:987-996. 10.1016/j.funbio.2011.04.002.

Hartley, G., and O'Neill, R.J. (2019). Centromere Repeats: Hidden Gems of the Genome. Genes (Basel) **10**10.3390/genes10030223.

Head, B.P., Patel, H.H., and Insel, P.A. (2014). Interaction of membrane/lipid rafts with the cytoskeleton: impact on signaling and function: membrane/lipid rafts, mediators of cytoskeletal arrangement and cell signaling. Biochim Biophys Acta **1838**:532-545. 10.1016/j.bbamem.2013.07.018.

Heckman, D.S., Geiser, D.M., Eidell, B.R., Stauffer, R.L., Kardos, N.L., and Hedges, S.B. (2001). Molecular evidence for the early colonization of land by fungi and plants. Science **293**:1129-1133. 10.1126/science.1061457.

Hedges, S.B., Blair, J.E., Venturi, M.L., and Shoe, J.L. (2004). A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evol Biol **4**:2. 10.1186/1471-2148-4-2.

Hernandez-Vasquez, M.N., Adame-Garcia, S.R., Hamoud, N., Chidiac, R., Reyes-Cruz, G., Gratton, J.P., Cote, J.F., and Vazquez-Prado, J. (2017). Cell adhesion controlled by adhesion G protein-coupled receptor GPR124/ADGRA2 is mediated by a protein complex comprising intersectins and Elmo-Dock. The Journal of biological chemistry **292**:12178-12191. 10.1074/jbc.M117.780304.

Herron, M.D., Hackett, J.D., Aylward, F.O., and Michod, R.E. (2009). Triassic origin and early radiation of multicellular volvocine algae. Proc Natl Acad Sci U S A **106**:3254-3258. 10.1073/pnas.0811205106.

Hesford, F.J., Berger, E.G., and Van den Eijnden, D.H. (1981). Identification of the product formed by human erythrocyte galactosyltransferase. Biochim Biophys Acta **659**:302-311. 10.1016/0005-2744(81)90056-5.

Hillmeier, M., Wagner, M., Ensfelder, T., Korytiakova, E., Thumbs, P., Muller, M., and Carell, T. (2021). Synthesis and structure elucidation of the human tRNA nucleoside mannosyl-queuosine. Nat Commun **12**:7123. 10.1038/s41467-021-27371-9.

Hoffmann, A., Erber, L., Betat, H., Stadler, P.F., Morl, M., and Fallmann, J. (2021). Changes of the tRNA Modification Pattern during the Development of Dictyostelium discoideum. Noncoding RNA **7**10.3390/ncrna7020032.

Horas, E.L., Theodosiou, L., and Becks, L. (2018). Why Are Algal Viruses Not Always Successful? Viruses **10**10.3390/v10090474.

Humphrey, J.D., Dufresne, E.R., and Schwartz, M.A. (2014). Mechanotransduction and extracellular matrix homeostasis. Nat Rev Mol Cell Biol **15**:802-812. 10.1038/nrm3896.

Ji, C., Wang, Y., Guo, X., Hartson, S., and Jiang, H. (2004). A pattern recognition serine proteinase triggers the prophenoloxidase activation cascade in the tobacco hornworm, Manduca sexta. The Journal of biological chemistry **279**:34101-34106. 10.1074/jbc.M404584200.

Kajikawa, S., Ezura, Y., Izu, Y., Nakashima, K., Noda, M., and Nifuji, A. (2022). Profilin-1 negatively controls osteoclast migration by suppressing the protrusive structures based on branched actin filaments. J Bone Miner Metab **40**:561-570. 10.1007/s00774-022-01320-y.

Klar, A., Baldassare, M., and Jessell, T.M. (1992). F-spondin: a gene expressed at high levels in the floor plate encodes a secreted protein that promotes neural cell adhesion and neurite extension. Cell **69**:95-110. 10.1016/0092-8674(92)90121-r.

Kloareg, B., Badis, Y., Cock, J.M., and Michel, G. (2021). Role and Evolution of the Extracellular Matrix in the Acquisition of Complex Multicellularity in Eukaryotes: A Macroalgal Perspective. Genes (Basel) **12**10.3390/genes12071059.

Ko, C.P. (1987). A lectin, peanut agglutinin, as a probe for the extracellular matrix in living neuromuscular junctions. J Neurocytol **16**:567-576. 10.1007/BF01668509.

Koselski, M., Pupkis, V., Hashimoto, K., Lapeikaite, I., Hanaka, A., Wasko, P., Plukaite, E., Kuchitsu, K., Kisnieriene, V., and Trebacz, K. (2021). Impact of Mammalian Two-Pore Channel Inhibitors on Long-Distance Electrical Signals in the Characean Macroalga Nitellopsis obtusa and the Early Terrestrial Liverwort Marchantia polymorpha. Plants (Basel) **10**10.3390/plants10040647.

Lang, D., Weiche, B., Timmerhaus, G., Richardt, S., Riano-Pachon, D.M., Correa, L.G., Reski, R., Mueller-Roeber, B., and Rensing, S.A. (2010). Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. Genome Biol Evol **2**:488-503. 10.1093/gbe/evq032.

Libert, F., and Vassart, G. (1989). Structure-function relationships of the complement components. Immunol Today **10**:407. 10.1016/0167-5699(89)90036-4.

Ma, R., Huang, B., Chen, J., Huang, Z., Yu, P., Ruan, S., and Zhang, Z. (2021). Genome-wide identification and expression analysis of dirigent-jacalin genes from plant chimeric lectins in Moso bamboo (Phyllostachys edulis). Plos One **16**:e0248318. 10.1371/journal.pone.0248318.

Macaisne, N., Liu, F., Scornet, D., Peters, A.F., Lipinska, A., Perrineau, M.M., Henry, A., Strittmatter, M., Coelho, S.M., and Cock, J.M. (2017). The Ectocarpus IMMEDIATE UPRIGHT gene encodes a member of a novel family of cysteine-rich proteins with an unusual distribution across the eukaryotes. Development **144**:409-418. 10.1242/dev.141523.

Mahmood, N., Moosa, M.M., Tamanna, N., Sarker, S.K., Najnin, R.A., and Alam, S.S. (2012). In silico analysis reveals the presence of a large number of Ankyrin repeat containing proteins in Ectocarpus siliculosus. Interdiscip Sci **4**:291-295. 10.1007/s12539-012-0134-9.

Moulessehoul, S., Hirsch, M., and Pouliquen, Y. (1992). Lectin-binding sites in quick-frozen, freeze-substituted extracellular matrix of rabbit corneal stroma: a comparative electron microscopic study with different embedding procedures. Exp Eye Res **54**:307-312. 10.1016/s0014-4835(05)80220-9.

Munakata, H., Nakada, T., Nakahigashi, K., Nozaki, H., and Tomita, M. (2016). Phylogenetic Position and Molecular Chronology of a Colonial Green Flagellate, Stephanosphaera pluvialis (Volvocales, Chlorophyceae), among Unicellular Algae. J Eukaryot Microbiol **63**:340-348. 10.1111/jeu.12283.

Nagafuchi, A., and Takeichi, M. (1988). Cell binding function of E-cadherin is regulated by the cytoplasmic domain. EMBO J **7**:3679-3684. 10.1002/j.1460-2075.1988.tb03249.x.

Nagasato, C., Yonamine, R., and Motomura, T. (2022). Ultrastructural Observation of Cytokinesis and Plasmodesmata Formation in Brown Algae. Methods Mol Biol **2382**:253-264. 10.1007/978-1-0716-1744-1_16.

Nayak, R.C., Chang, K.H., Vaitinadin, N.S., and Cancelas, J.A. (2013). Rho GTPases control specific cytoskeleton-dependent functions of hematopoietic stem cells. Immunol Rev **256**:255-268. 10.1111/imr.12119.

Nelson, D.R., Hazzouri, K.M., Lauersen, K.J., Jaiswal, A., Chaiboonchoe, A., Mystikou, A., Fu, W., Daakour, S., Dohai, B., Alzahmi, A., et al. (2021). Large-scale genome sequencing reveals the driving forces of viruses in microalgal evolution. Cell Host Microbe **29**:250-266 e258. 10.1016/j.chom.2020.12.005.

Neu, T.R., and Kuhlicke, U. (2017). Fluorescence Lectin Bar-Coding of Glycoconjugates in the Extracellular Matrix of Biofilm and Bioaggregate Forming Microorganisms. Microorganisms **5**10.3390/microorganisms5010005.

Outenreath, R.L., Roberson, M.M., and Barondes, S.H. (1988). Endogenous lectin secretion into the extracellular matrix of early embryos of Xenopus laevis. Dev Biol **125**:187-194. 10.1016/0012-1606(88)90071-1.

Ozhogina, O.A., Trexler, M., Banyai, L., Llinas, M., and Patthy, L. (2001). Origin of fibronectin type II (FN2) modules: structural analyses of distantly-related members of the kringle family idey the kringle domain of neurotrypsin as a potential link between FN2 domains and kringles. Protein Sci **10**:2114-2122. 10.1110/ps.15801.

Parfrey, L.W., Lahr, D.J., Knoll, A.H., and Katz, L.A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc Natl Acad Sci U S A **108**:13624-13629. 10.1073/pnas.1110633108.

Paulson, A.F., Mooney, E., Fang, X., Ji, H., and McCrea, P.D. (2000). Xarvcf, Xenopus member of the p120 catenin subfamily associating with cadherin juxtamembrane region. The Journal of biological chemistry **275**:30124-30131. 10.1074/jbc.M003048200.

Peters, A.F., Scornet, D., Ratin, M., Charrier, B., Monnier, A., Merrien, Y., Corre, E., Coelho, S.M., and Cock, J.M. (2008). Life-cycle-generation-specific developmental processes are modified in the immediate upright mutant of the brown alga Ectocarpus siliculosus. Development **135**:1503-1512. 10.1242/dev.016303.

Ratcliff, W.C., Denison, R.F., Borrello, M., and Travisano, M. (2012). Experimental evolution of multicellularity. Proc Natl Acad Sci U S A **109**:1595-1600. 10.1073/pnas.1115323109.

Reid, K.B., and Day, A.J. (1989). Structure-function relationships of the complement components. Immunol Today **10**:177-180. 10.1016/0167-5699(89)90317-4.

Ricci-Azevedo, R., Roque-Barreira, M.C., and Gay, N.J. (2017). Targeting and Recognition of Toll-Like Receptors by Plant and Pathogen Lectins. Front Immunol **8**:1820. 10.3389/fimmu.2017.01820.

Schallus, T., Jaeckh, C., Feher, K., Palma, A.S., Liu, Y., Simpson, J.C., Mackeen, M., Stier, G., Gibson, T.J., Feizi, T., et al. (2008). Malectin: a novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein N-glycosylation. Mol Biol Cell **19**:3404-3414. 10.1091/mbc.e08-04-0354.

Schwarz, E., and Benzer, S. (1999). The recently reported NIbeta domain is already known as the Calx-beta motif. Trends Biochem Sci **24**:260. 10.1016/s0968-0004(99)01422-x.

Schwarz, E.M., and Benzer, S. (1997). Calx, a Na-Ca exchanger gene of Drosophila melanogaster. Proc Natl Acad Sci U S A **94**:10249-10254. 10.1073/pnas.94.19.10249.

Singh, B., Maharjan, S., Kim, Y.K., Jiang, T., Islam, M.A., Kang, S.K., Cho, M.H., Choi, Y.J., and Cho, C.S. (2014). Targeted gene delivery via N-acetylglucosamine receptor mediated endocytosis. J Nanosci Nanotechnol **14**:8356-8364. 10.1166/jnn.2014.9919.

Song, G., Lazar, G.A., Kortemme, T., Shimaoka, M., Desjarlais, J.R., Baker, D., and Springer, T.A. (2006). Rational design of intercellular adhesion molecule-1 (ICAM-1) variants for antagonizing integrin lymphocyte function-associated antigen-1-dependent adhesion. The Journal of biological chemistry **281**:5042-5049. 10.1074/jbc.M510454200.

Takada, Y. (2012). Potential role of kringle-integrin interaction in plasmin and uPA actions (a hypothesis). J Biomed Biotechnol **2012**:136302. 10.1155/2012/136302.

Tang, X.Y., Li, Y.F., and Tan, S.M. (2008). Intercellular adhesion molecule-3 binding of integrin alphaL beta2 requires both extension and opening of the integrin headpiece. J Immunol **180**:4793-4804. 10.4049/jimmunol.180.7.4793.

Terauchi, M., Nagasato, C., and Motomura, T. (2015). Plasmodesmata of brown algae. J Plant Res **128**:7-15. 10.1007/s10265-014-0677-4.

Thompson, C.J., Vu, V.H., Leckband, D.E., and Schwartz, D.K. (2021). Cadherin cis and trans interactions are mutually cooperative. Proc Natl Acad Sci U S A **118**10.1073/pnas.2019845118.

Tong, C.F., Zhang, Y., Lu, S.Q., Li, N., Gong, Y.X., Yang, H., Feng, S.L., Du, Y., Huang, D.D., and Long, M. (2018). Binding of intercellular adhesion molecule 1 to beta2-integrin regulates distinct cell adhesion processes on hepatic and cerebral endothelium. Am J Physiol Cell Physiol **315**:C409-C421. 10.1152/ajpcell.00083.2017.

Van Lanen, S.G., Reader, J.S., Swairjo, M.A., de Crecy-Lagard, V., Lee, B., and Iwata-Reuyl, D. (2005). From cyclohydrolase to oxidoreductase: discovery of nitrile reductase activity in a common fold. Proc Natl Acad Sci U S A **102**:4264-4269. 10.1073/pnas.0408056102.

Wei, X., Orchardson, M., Gracie, J.A., Leung, B.P., Gao, B., Guan, H., Niedbala, W., Paterson, G.K., McInnes, I.B., and Liew, F.Y. (2001). The Sushi domain of soluble IL-15 receptor alpha is essential for binding IL-15 and inhibiting inflammatory and allogenic responses in vitro and in vivo. J Immunol **167**:277-282. 10.4049/jimmunol.167.1.277.

Willoughby, L.F., Manent, J., Allan, K., Lee, H., Portela, M., Wiede, F., Warr, C., Meng, T.C., Tiganis, T., and Richardson, H.E. (2017). Differential regulation of protein tyrosine kinase signalling by Dock and the PTP61F variants. The FEBS journal **284**:2231-2250. 10.1111/febs.14118.

Winograd-Katz, S.E., Fassler, R., Geiger, B., and Legate, K.R. (2014). The integrin adhesome: from genes and proteins to human disease. Nat Rev Mol Cell Biol **15**:273-288. 10.1038/nrm3769.

Yang, E.C., Boo, S.M., Bhattacharya, D., Saunders, G.W., Knoll, A.H., Fredericq, S., Graf, L., and Yoon, H.S. (2016). Divergence time estimates and the evolution of major lineages in the florideophyte red algae. Sci Rep **6**:21361. 10.1038/srep21361.

Yap, A.S., Niessen, C.M., and Gumbiner, B.M. (1998). The juxtamembrane region of the cadherin cytoplasmic tail supports lateral clustering, adhesive strengthening, and interaction with p120ctn. J Cell Biol **141**:779-789. 10.1083/jcb.141.3.779.

Yonemura, S., Hirao, M., Doi, Y., Takahashi, N., Kondo, T., Tsukita, S., and Tsukita, S. (1998). Ezrin/Radixin/Moesin (ERM) Proteins Bind to a Positively Charged Amino Acid Cluster in the Juxta-Membrane Cytoplasmic Domain of CD44, CD43, and ICAM-2. Journal of Cell Biology **140**:885-895. 10.1083/jcb.140.4.885.

Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G., and Bhattacharya, D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol **21**:809-818. 10.1093/molbev/msh075.

Zaidel-Bar, R., and Geiger, B. (2010). The switchable integrin adhesome. J Cell Sci **123**:1385-1388. 10.1242/jcs.066183.

Zaidel-Bar, R., Itzkovitz, S., Ma'ayan, A., Iyengar, R., and Geiger, B. (2007). Functional atlas of the integrin adhesome. Nat Cell Biol **9**:858-867. 10.1038/ncb0807-858.

Zarbock, A., Lowell, C.A., and Ley, K. (2007). Spleen tyrosine kinase Syk is necessary for E-selectin-induced alpha(L)beta(2) integrin-mediated rolling on intercellular adhesion molecule-1. Immunity **26**:773-783. 10.1016/j.immuni.2007.04.011.

Zhang, W., and Pan, T. (2021). Quantitative probing of glycosylated queuosine modifications in tRNA. Methods Enzymol **658**:73-82. 10.1016/bs.mie.2021.06.003.

Zheng, Y., and Leftheris, K. (2020). Insights into Protein-Ligand Interactions in Integrin Complexes: Advances in Structure Determinations. J Med Chem **63**:5675-5696. 10.1021/acs.jmedchem.9b01869.

Zimmer, A., Lang, D., Richardt, S., Frank, W., Reski, R., and Rensing, S.A. (2007). Dating the early evolution of plants: detection and molecular clock analyses of orthologs. Mol Genet Genomics **278**:393-402. 10.1007/s00438-007-0257-6.

Zisman, S., Marom, K., Avraham, O., Rinsky-Halivni, L., Gai, U., Kligun, G., Tzarfaty-Majar, V., Suzuki, T., and Klar, A. (2007). Proteolysis and membrane capture of F-spondin generates combinatorial guidance cues from a single molecule. J Cell Biol **178**:1237-1249. 10.1083/jcb.200702184.