Insulated Outlier Chromosomes Drive Metabolic and Evolutionary Innovation in Minimal Eukaryotic Algae

Martha Valiadi[1,2,*], Keith Harrison[1,3,*], Yann Loe-Mie[4], Bryony Williams[3], Dyan Ankrett[3], Nicholas Smirnoff[3] & Adam Monier[1,3]

[1]Living Systems Institute, University of Exeter, Exeter, UK; [2]Foundation for Research and Technology Hellas, Institute of Molecular Biology and Biotechnology, Heraklion, Crete, Greece; [3]Biosciences, Faculty of Health and Life Sciences, University of Exeter, Exeter, UK; [4]Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics Hub, Paris, France.

*Equal contributions

Correspondence: a.monier@exeter.ac.uk

Abstract

Mamiellales, a group of marine picoeukaryotes including *Ostreococcus*, the smallest known free-living eukaryote, exhibit ecological success despite compact genomes. We investigate how these organisms retain adaptive potential under such constraints, focusing on two outlier chromosomes with extensive low-GC regions. Using an integrated multi-omics approach—Hi-C, transcriptomics, genome-wide methylation analysis, and metabolomics—we show these atypical regions act as genomic innovation hubs, enriched in transposable elements, introns, and highly expressed genes. These regions display reduced methylation and are spatially insulated by the only topologically associating domain in the genome. Notably, they harbour giant type I polyketide synthase genes, whose diversity across *Ostreococcus* species is mirrored by distinct polyketide profiles. We propose that Mamiellales localise evolutionary experimentation to these insulated regions, balancing genome minimisation with adaptive flexibility. Our findings reveal a novel strategy for maintaining genomic plasticity in streamlined eukaryotes and provide new insight into the mechanisms driving Mamiellales' ecological success.

Introduction

The marine green alga *Ostreococcus* is the smallest known free-living eukaryote (~0.8 µm), with a single plastid and mitochondrion[1,2]. Its minimalistic cellular organisation presents a simplified eukaryotic model for studying fundamental processes such as photosynthesis, circadian rhythms, and environmental responses[3–5]. *Ostreococcus* belongs to Mamiellales (class Mamiellophyceae), which includes other picoalgae like *Bathycoccus*, *Mantoniella*, and *Micromonas*[6]. Mamiellales thrive globally, from tropical to polar waters[7–9], supported by ecotype differentiation and intra-order genetic diversity[10–12]. This suggests a strong adaptive capacity, despite—or because of—their extensive genomic streamlining. For example, *O. tauri* has a 12.9 Mb genome encoding ~8,000 genes, which is much smaller and denser than *Chlamydomonas reinhardtii*[13]. Their

small, haploid, gene-dense genomes[14–19] confer efficiency under resource limitation[20], yet how these genomes retain sufficient adaptive capacity remains unclear[21].

A hallmark feature of Mamiellales genomes may hold the answer: two outlier chromosomes—the so-called big and small outlier chromosomes (BOC and SOC)—each containing regions of markedly lower GC content (BOC1 and SOC1) [22]. These atypical regions are enriched in introns, transposable elements (TEs), and highly expressed genes[23–25], and vary significantly among *Ostreococcus* strains[26]. These polymorphisms are associated with phenotypic differences in growth and viral susceptibility[26,27]. In *O. tauri*, the BOC includes a mating-type locus, and genes linked to chromatin conformation[23,26,28], while the SOC, highly variable due to duplication and translocation, is involved in antiviral defence[27,29].

We hypothesise that BOC1 and SOC1 act as insulated hotspots for genomic innovation, enabling adaptation without compromising genome compactness. Their TE and intron richness may promote rearrangement and splicing diversity[30,31], while low GC content may affect chromatin structure and gene regulation[32–34]. Elevated gene expression in these regions suggests a distinct regulatory environment[23,24,35]. To test this, we employed a multi-omics approach—Hi-C, transcriptomics, genome-wide methylation analysis, and metabolomics—to examine the structure, regulation, and functional output of the outlier chromosomes. We provide evidence that picoalgae compartmentalise adaptive functions to resolve the tension between genome minimisation and evolutionary responsiveness.

Results

Structural conservation and functional specialisation of outlier regions across Mamiellales

To assess the conservation of outlier regions across Mamiellales, we analysed BOC1 and SOC1 sequences from six species with publicly available genome assemblies: *O. tauri*, *Ostreococcus lucimarinus*, *Ostreococcus sp.* RCC809, *Micromonas commoda*, *Micromonas pusilla*, and *Bathycoccus prasinos*[15,16,21,23]. Despite their minimalistic genomes (average size: 16.8 ± 4.4 Mb, 533 ± 60.7 genes per Mb; Supplementary Tables 1, 2), these low-GC outlier regions are conserved. Further evidence came from *Mantoniella tinhauana*[36], where outlier regions had been questioned. We identified two low-GC regions (~15% lower GC) in its genome assembly (Supplementary Fig. 1; Supplementary Table 2). Although BOC and SOC chromosomes in *M. tinhauana* are larger (~400–500 kb), their BOC1 and SOC1 regions are similar in size to those in other Mamiellales. This consistency suggests these regions are ancient and conserved across the order.

We conducted protein sequence similarity searches to examine variations in gene content and organisation. While SOC1 regions showed limited synteny and fewer orthologues across genera (Supplementary Fig. 2), BOC1 regions exhibited greater conservation, though not strict synteny (Fig. 1a). For example, the *O. tauri* BOC1 shares 120 orthologues with *Ostreococcus sp.* RCC809 and 156 with *O. lucimarinus*, and 99 orthologues with *B. prasinos*. In *Micromonas* species, BOC1 regions are larger (1.55–1.67 Mb) and share fewer orthologues with *Ostreococcus*.

Gene Ontology (GO) enrichment analyses on *O. tauri* BOC1 revealed an abundance of genes related to DNA transposition, transcription, and integration (Fig. 1b; Extended Data 1), indicating high TE activity and genomic plasticity. It also contains genes involved in chromatin remodelling, including mini-chromosome maintenance (MCM) complex components, nucleolar proteins, and sister chromatid cohesion protein 2 (SCC2). The SOC1 region, in contrast, is enriched in genes related to RNA polymerase activity, membrane components, and methylation (Extended Data 1).

Transposable elements and introns cluster in outlier genomic regions across Mamiellales genomes

If outlier regions serve as hubs for genomic innovation, they should contain higher concentrations of elements driving genomic rearrangements and genetic diversity. We conducted a survey of introns and TEs across Mamiellales genomes, confirming previous findings[14,36,37] and revealing a distinct, non-random distribution of these features. Intron density is notably higher in BOC1 regions across all taxa, except *M. tinhauana* (Extended Data 2). For example, in *O. tauri* RCC4221, intron density is four times higher in BOC1 than in the standard genome.

The enrichment of TEs in outlier regions is even more pronounced. BOC1 and SOC1 in *O. tauri* RCC4221 contain 14.5 and 17.4 times more TEs, respectively, compared to the standard genome (Fig. 2a, b; Supplementary Fig. 3; Extended Data 2). This pattern is consistent across most Mamiellales species, except *B. prasinos* and *O. lucimarinus*. In *M. tinhauana*, both low-GC regions also show significant TE enrichment (Supplementary Fig. 1; Extended Data 2), confirming their classification as BOC1 and SOC1. TE densities vary between species, with *M. pusilla* CCMP1545 having the highest concentration, with 738 TEs/Mb in BOC1 and 484 TEs/Mb in SOC1, followed by *M. tinhauana* RCC11003.

Various TE classes dominate these regions (Fig. 2a; Extended Data 2). LTR retrotransposons (Ty1/Copia, Ty3/Gypsy), DNA transposons (Tc1/Mariner, PiggyBac, Harbinger), and Mutator, hAT, and Helitrons are most prevalent. Ty1/Copia elements are especially concentrated in BOC1 and SOC1. In *O. tauri* RCC4221, Ty1/Copia

densities rise from 2.04 elements per Mb in the standard genome to 37.4 and 167 elements per Mb in BOC1 and SOC1, respectively (Extended Data 2).

TEs often reside near actively transcribed genes. In *O. tauri*, transcriptome analysis revealed associations between TEs and transcriptionally active genes in BOC1 and SOC1 (Fig. 2c-f; Extended Data 2). A ~5 kb Ty1/Copia retrotransposon is adjacent to a hexosyltransferase gene (Fig. 2c), while a ~1 kb Tc1/Mariner DNA transposon is near a transcription factor IIA gene (Fig. 2d). Mutator-like elements (MULEs) are frequently found near PKS genes with repeated domains, and some MULEs are inserted directly within PKS coding sequences (Fig. 2e, f). These PKS genes show variable transcription levels, with some domains highly expressed and others showing low or undetectable expression.

Genes on outlier genomic regions are more highly expressed and differentially spliced

To assess whether genes in outlier regions are expressed differently, we analysed transcriptomic data from O. tauri RCC4221 and other Mamiellales species. Genes in BOC1 regions show consistently higher expression compared to the standard genome across all species. In *O. tauri*, BOC1 genes exhibit a 3.8-fold higher expression (Kruskal-Wallis, $p<0.001$; Fig. 3a, e), with *M. commoda*, *M. pusilla*, and *B. prasinos* showing similar increases (2.4, 3.3, and 1.6-fold, respectively; $p<0.001$).

A negative correlation between GC content and gene expression was found in *O. tauri*, *M. commoda*, and *M. pusilla* (Spearman's rank, $\rho=-0.37$, -0.29, and -0.49, $p<0.001$; Fig. 3i, j, k), while *B. prasinos* showed the opposite pattern ($\rho=0.22$, $p<0.001$; Fig. 3l). This suggests genus-specific regulatory mechanisms.

Additionally, outlier region genes show substantial alternative splicing. In O. tauri, 38% (1,216 of 3,215) of splice variants come from BOC1 or SOC1 regions, despite these regions housing fewer genes. A similar trend is observed in *O. lucimarinus*, with 20% of splice variants mapping to outlier regions (Extended Data 2).

Three-dimensional genome architecture reveals structural isolation of outlier regions

We performed chromosome Hi-C analysis on the *O. tauri* RCC4221 genome to investigate the 3D organisation of outlier regions. This analysis revealed structural features that insulate these regions from the rest of the genome. Our DNA-DNA contact map (Fig. 4a; Supplementary Fig. 4), constructed from 330 million contacts at 10 kb resolution, shows a simple chromosomal architecture dominated by intrachromosomal interactions (77% of all contacts; Extended Data 3). Both outlier chromosomes display enhanced intrachromosomal contacts,

together accounting for ~20% of all intrachromosomal interactions. These enhanced contacts appear as higher frequencies extending beyond the diagonal in Hi-C maps (Fig. 4; Supplementary Fig. 5).

The BOC region exhibits a distinct organisation, with a topologically associated domain (TAD)-like structure encompassing BOC1 (Fig. 4b, c; Supplementary Fig. 6; Extended Data 3). Within this domain, we identified four putative chromatin loops (α, β, γ, δ). Genes in these loops are involved in diverse functions, such as glycolipid biosynthesis and chromatin binding (Extended Data 3). The BOC1 region shows a unique interaction pattern, indicating spatial compartmentalisation within chromosome 2.

The SOC region also displays a distinctive interaction pattern (Fig. 4d). Its Hi-C map suggests a TAD-like structure encompassing SOC1, though its size limits domain identification. Quantitative analysis confirmed that BOC and SOC have fewer interchromosomal interactions compared to standard genome regions (Fig. 4e; $p<0.0001$; Extended Data 3). Instead, these outlier regions preferentially interact with each other, suggesting regulatory isolation.

Finally, we examined Hi-C data from *M. tinhauana*[36], and found similar structural isolation of outlier regions (Supplementary Fig. 1; Extended Data 3), further supporting the conservation of these genomic features.

Outlier genomic regions maintain distinctive epigenetic landscapes

To assess whether structurally isolated regions also exhibit unique epigenetic features, we examined DNA methylation across the *O. tauri* RCC4221 genome. This revealed striking patterns that further distinguish outlier regions. A sharp shift in methylation at ~0.55 Mb on chromosome 2 precisely demarcates the BOC1 region (Fig. 4b), aligning with the TAD-like boundary identified by Hi-C, as well as transitions in transposable element (TE) density and gene expression. This convergence of structural, regulatory, and epigenetic features suggests coordinated regulation at the BOC1 boundary.

Methylation levels within outlier regions are consistently lower than in the standard genome. BOC1 shows a 1.4-fold reduction in CpG 5-methylcytosine (5mCG) relative to the rest of the genome (Fig. 4b; Supplementary Figs. 8, 9; Extended Data 4). Overall, 5-methylcytosine (5mC) and N6-methyladenine (6mA) are reduced by 1.7- and 1.4-fold, respectively, while 5-hydroxymethylcytosine (5hmC) and 5hmCG remain uniformly low across the chromosome. SOC1 also displays distinctive methylation, with a sharp rise in 6mA at its start and increased 5mCG at its end (Supplementary Fig. 10), reinforcing its identity as a separate compartment.

To test whether these methylation patterns are conserved across Mamiellales, we analysed Bisulfite-seq data from *O. lucimarinus* and *M. pusilla*[38]. In *O. lucimarinus*, CpG methylation drops from 22.2% in standard regions to 19.2% in BOC1 and 16.2% in SOC1. *M. pusilla* shows a similar trend, with levels decreasing from 17.8% to

14.5% in BOC1 and 10.9% in SOC1 (Fig. 5a,b; Supplementary Fig. 11; Extended Data 4). These conserved patterns highlight the distinct epigenetic signature of Mamiellales outlier regions.

Nucleosome positioning and TE associations

To further characterise the chromatin landscape of outlier regions, we analysed nucleosome positioning and its relationship to gene expression using public MNase-Seq and RNA-Seq data from *O. lucimarinus* and *M. pusilla*. Outlier regions displayed distinct chromatin features.

Consistent with earlier observations (Fig. 3), RNA-Seq coverage in BOC1 was significantly higher than in standard regions: 1.86× in *O. lucimarinus* and 2.44× in *M. pusilla* (Kruskal–Wallis, $p < 0.001$). This increase correlated with distinctive nucleosome features: higher occupancy at transcription start sites (TSS), less defined nucleosome centres in 5' UTRs, and lower CpG methylation (Fig. 5c,e), suggesting a more open chromatin state conducive to transcription. Additionally, TEs remain enriched in the BOC1 of *M. pusilla* (3.45×), a region notably lacking introner elements (IEs) despite their abundance in the standard genome[16,39] (Fig. 5b).

Autocorrelation analysis revealed tighter and more variable nucleosome spacing in BOC1. In *O. lucimarinus*, spacing decreased from 183 to 181 bp, and in *M. pusilla*, from 207 to 196 bp (Fig. 5d,f). These patterns were specific to outlier regions, confirmed by analysis of standard chromosomes 1 and 3 (Supplementary Fig. 12). Nucleosome centres were more well positioned in BOC1 and SOC1 (Supplementary Fig. 12e,f; Extended Data 4), and stiffness was slightly reduced compared to standard regions (Extended Data 4).

We also examined TE–nucleosome interactions. In *M. pusilla*, TEs showed a bimodal distribution with peaks ±50 bp from nucleosome centres, more prominent in BOC1 (Supplementary Fig. 13a,b). In *O. lucimarinus*, TEs peaked at nucleosome centres genome-wide, but BOC1-specific patterns couldn't be assessed due to low TE abundance. These chromatin patterns mirror those observed in *Arabidopsis thaliana*, where the histone variant H2A.Z plays a critical role in driving TE insertions and modulating chromatin accessibility. Through phylogenetic analysis, we identified H2A.Z variants in Mamiellales taxa that are closely related to plant H2A.Z histone variants (Supplementary Fig. 14).

Notably, regions encoding giant type I PKS genes within BOC1 showed elevated MNase-Seq coverage and CpG methylation (Fig. 5a,b; Extended Data 4), suggesting targeted regulation of these genes, which are involved in secondary metabolism and may contribute to environmental adaptation.

BOC1 encodes diverse polyketide biosynthetic pathways with unique metabolic outputs

To address whether genomic innovation in outlier regions confers functional diversity, we investigated giant type I PKS genes within BOC1—hallmarks of functional novelty within these specialised compartments. Our central question was: Do structural variations in these PKS genes translate to measurable differences in polyketide production across Mamiellales species?

BOC1 regions across Mamiellales contain large, modular PKS genes with extensive variation in size and architecture (Fig. 6a; Extended Data 5). Gene lengths range from 14,393 bp in *O. tauri* to 64,971 bp in *M. tinhauana*, driven by differences in repeated modules. Intra-species variability is also pronounced, with *O. tauri* strains showing PKS genes between 14,393 and 33,947 bp.

To characterise this variation in our laboratory subclone of *O. tauri* RCC4221 ('Exeter'), we re-sequenced its genome using nanopore long reads. While BOC1 (~550 kb) and SOC1 (~150 kb) retained similar size and location compared to the reference *O. tauri* strain, the PKS gene in 'Exeter' was substantially larger and structurally distinct from the published reference[21,40]. We observed numerous inversions in the outlier chromosomes that were absent in standard chromosomes (Supplementary Fig. 15). TE types matched those in the reference strain but were more abundant (Extended Data 2). The reference PKS gene encodes repeated ketosynthase (KS), ketoreductase (KR), adenylate-forming domain (AFD), and acyl carrier protein (ACP) domains, while the 'Exeter' variant exhibits an expanded architecture with additional modules (Fig. 6a).

To determine whether these structural variations yield distinct metabolic profiles, we performed high-resolution mass spectrometry on two *Ostreococcus* species, *O. tauri* ('Exeter') and *O. lucimarinus* (CCE9901), and systematically mined the data for candidate polyketides based on predicted structures from the PKS gene architectures. This approach revealed diverse metabolite peaks in both *Ostreococcus* species, several matching predicted polyketide masses and structures (Fig. 6b).

In both species, we detected a series of compounds that align with the predicted products of the 11,265 and 18,194 aa giant type I PKS proteins present in their respective genomes. The major polyketide products were shared between organisms, with products similar in m/z to mycolactone, ebelactone A, 6-deoxyerythronolide B, and actinoallolide A being abundant in both species (Fig. 6b). Mass spectrometry analysis revealed mycolactone ([M+K-2H]-, m/z 742.502) had a mean abundance of 38,127.26 ± 11,043.75 in *O. tauri* and 19,340.36 ± 16,228.74 in *O. lucimarinus*. Actinoallolide A ([M+CH3OH+H]+, m/z 564.3662) showed higher abundance in *O. tauri* (20,867.13 ± 11,860.27) compared to *O. lucimarinus* (7,475.06 ± 4,234.36). 6-deoxyerythronolide B ([M+H-2H2O]+, m/z 386.2668) was detected at mean values of 24,590.81 ± 2,293.26 in *O. tauri* and 14,965.08 ± 13,869.67 in *O. lucimarinus*. Notably, ebelactone A was detected with two different adducts: as [M-H2O-H]- (m/z 370.2719) with mean abundances of 21,710.16 ± 4,967.15 in *O. tauri* and 36,676.53 ± 28,232.34 in *O.*

lucimarinus, and as [M+H-2H2O]+ (m/z 370.2719) with significantly higher abundance in *O. tauri* (60,867.73 ± 5,678.97) compared to *O. lucimarinus* (13,612.24 ± 6,164.75).

Outlier regions with similar features identified in phylogenetically distant picoalgae

To explore whether specialised genomic regions are unique to Mamiellales or represent a broader evolutionary strategy, we examined genomic data from picoeukaryotes. Remarkably, we identified similar outlier regions in the distantly related *Pelagomonas calceolata*[41] stramenopile picoalga (Supplementary Fig. 16). Each chromosome in *P. calceolata* contains an outlier region (~500 kb) with ~15% lower GC content compared to the rest of the genome, mirroring the pattern observed in Mamiellales. Hi-C mapping further revealed that these regions are structurally insulated from the standard genome, consistent with the compartmentalisation seen in Mamiellales BOC1 and SOC1.

One of these outlier regions contains a giant PKS gene adjacent to a MULE TE (Supplementary Fig. 17), strikingly similar to the genomic arrangement in Mamiellales BOC1. Analysis of available transcriptomic data[41] demonstrated that genes within these outlier regions are expressed at ~3-fold higher levels than those in standard genomic regions. Our reanalysis of published experimental RNA-Seq data from nitrogen limitation experiments[42] revealed that *P. calceolata* PKS genes in GC outlier regions respond differentially to nitrogen availability. The non-modular PKS (Pca_2p18890) is downregulated under moderate nitrogen limitation (441 vs 882 µM nitrate), while two components of a giant type 1 PKS cluster (Pca_2p00460, Pca_2p00430) show similar downregulation at 441 µM nitrate. Intriguingly, a third component (Pca_2p00540) is upregulated under more severe nitrogen limitation (220 µM) (all genes, $p < 0.05$, $|LFC| > 1$).

Discussion

Our findings reveal the significance of low-GC outlier regions as innovation hubs that reconcile streamlined genome architectures with the capacity for rapid adaptation in dynamic marine environments. By combining Hi-C, transcriptomics, DNA methylation profiling, and metabolomics in *O. tauri*, we uncover a complex interplay between genome organisation, epigenetic regulation and TE dynamics that contribute to the unique genomic landscape of this picoeukaryote. Extending our analyses to the wider Mamiellales and a stramenopile species, we show how this mechanism may drive evolutionary innovation in eukaryotic pico-algae, challenging their apparent genomic and structural simplicity[2,14,15,43].

Unique features of the BOC1 and SOC1 regions

Genes in BOC1 and SOC1 regions consistently show higher expression levels than their counterparts in the standard genome, a pattern conserved across all Mamiellales species we examined. This pattern in gene expression aligns with observations in land plants such as *Oryza sativa* and *Zea mays*[44], where low-GC content

often correlates with increased transcription, but contrasts with findings in the freshwater green alga *C. reinhardtii*[34]. Integrated analysis of MNase-seq and DNA methylation data revealed distinct epigenetic configurations in BOC1 and SOC1, including reduced levels of 5mCG methylation typically associated with open chromatin states[45–47]. These permissive epigenetic landscapes likely underpin the high transcriptional output in these outlier regions.

Beyond transcriptional enhancement, BOC1 and SOC1 also show dynamic nucleosome positioning. Nucleosome centres are more well-defined in BOC1 and SOC1 compared to standard genomic regions, with tighter and more variable spacing. In *O. lucimarinus*, spacing decreased from 183 to 181 bp, and in *M. pusilla*, from 207 to 196 bp. This configuration may facilitate both high gene expression and local TE mobilisation[48,49]. The well-defined yet more closely spaced nucleosomes suggest a regulated nucleosome arrangement that can still provide access to transcriptional machinery and TE insertion sites[48,49]. The bimodal distribution of TEs around nucleosome centres in *M. pusilla* BOC1 further supports the role of chromatin organisation in governing TE activity. This pattern recalls that observed in *A. thaliana*, where the histone variant H2A.Z preferentially directs TE insertions toward stress-responsive genes while buffering genomic regions harbouring essential housekeeping genes from disruptive integration events[50]. Given that *O. tauri* and other Mamiellales taxa encode an H2A.Z homolog related to plant H2A.Z variants, a parallel mechanism may confine TEs to the BOC1 and SOC1 compartments, allowing exapted TEs to fuel adaptive innovation while minimising deleterious effects on core genomic functions.

We also identified an enrichment of alternative splicing events within BOC1 and SOC1, suggesting that these regions contribute to increased transcriptional diversity. This enrichment in alternative splicing is consistent with previous findings in *M. pusilla* and *M. commoda*[39]. In compact genomes like those of Mamiellales, alternative splicing may provide a mechanism for expanding proteomic diversity without requiring genomic expansion via gene duplication. The interplay between TE activity, chromatin dynamics, and splicing regulation within the outlier regions could further enhance the functional versatility of genes housed in BOC1 and SOC1, particularly those involved in secondary metabolism and environmental responses.

Our Hi-C analyses further underscore this model, revealing TAD-like structures in BOC1 and SOC1 that help insulate these outlier regions from the standard genome. Such compartmentalisation likely contains TE activity while coordinating expression of genes integral to specialised pathways, including polyketide biosynthesis. These observations support previous reports on *O. tauri*, where the SOC was found to be hypervariable and critical for antiviral defence[26,27].

The role of transposable elements in the evolution of polyketide synthases

TEs are established drivers of genomic innovation, facilitating gene duplication, domain shuffling and large-scale rearrangements[51]. In compact haploid Mamiellales genomes, which possess limited recombination[37], TEs may provide an alternative pathway for generating structural and functional diversity. We observed a recurring association between MULEs and PKS genes in *O. tauri* and other Mamiellales taxa. As MULEs can promote genomic rearrangements and gene duplication[52], their proximity to PKS clusters likely accelerates the diversification and combinatorial diversity of PKS modules, thereby expanding the repertoire of polyketides and conferring adaptative advantages under varying environmental conditions[40,53,54]. Indeed, genome resequencing of our *O. tauri* laboratory strain showed that module rearrangements underpin the distinctive PKS architectures and polyketide metabolites observed here. This pattern is consistent with earlier results on the intraspecific variation in PKS gene clusters in *O. tauri*[40], including natural deletions in some strains, hinting at strong selective pressures tied to microbial interactions or predation.

This TE-mediated process is not restricted to Mamiellales. In the pico-sized stramenopile *Pelagomonas*, we identified low-GC outlier loci with increased expression, TE insertions and evidence of PKS diversification, suggesting a broader occurrence of TE-driven metabolic innovation among marine eukaryotic phytoplankton. PKS enzymes are intrinsically diverse and often critical for survival and adaptation[55,56]. In the green alga *C. reinhardtii*, a giant type I PKS is crucial for sexual reproduction and stress tolerance[55], while in the haptophyte alga *Prymnesium parvum*, the massive "PKZILLA" enzymes mediate toxin production and ecological competitiveness[56]. By enabling genomic rearrangements that alter PKS module composition, TEs facilitate combinatorial diversity in secondary metabolism, potentially enhancing responses to environmental challenges.

Our integrated genomic and metabolic analyses provide evidence linking PKS gene architecture to specific polyketide production. The differential abundance of compounds like mycolactone, ebelactone A, and actinoallolide A between *O. tauri* and *O. lucimarinus* correlates with their distinct PKS gene structures. These metabolites likely serve diverse ecological functions, from predator deterrence to inter-species signalling[55,56], contributing to the remarkable ecological success of these minimal eukaryotes.

A Unifying model for metabolic innovation in minimal eukaryotic genomes

Although direct evidence of sexual reproduction in Mamiellales remains elusive, several genomic features suggest it may occur under certain conditions and shape genome architecture[28,57]. In *O. tauri*, the BOC1 region combines suppressed recombination with abundant TE content, which are hallmarks of mating-type loci in other eukaryotes[58]. By contrast, the BOC1 in *B. prasinos* harbours fewer TEs and lacks PKS genes, mirroring PKS deletions reported in some *O. tauri* strains[40].

The lower overlap in orthologous genes between *Ostreococcus* and *Micromonas* BOC1 regions suggests lineage-specific rearrangements or differential gene loss and gain events. This genetic divergence likely underpins distinct adaptations to different ecological niches. The structural insulation of these regions through TAD-like boundaries may serve to contain potentially disruptive TE activity while still enabling genomic innovation within these compartments. This insulation effect is further evidenced by our observation in *M. pusilla* where IEs are excluded from outlier regions, despite being widespread across its standard genome[39,59], suggesting that TAD-like structures may act as barriers preventing IE colonization or facilitating their selective removal from these regions.

In many eukaryotes, genomic regions displaying TE accumulation, distinct chromatin compartmentalisation, low-GC content, and low recombination rates often point to centromeric regions or, in the case of fungi, repeat-induced point mutation (RIP). However, these explanations do not align with our results in *Ostreococcus*. Centromeres typically adopt heterochromatic states with marked transcriptional repression[60], while RIPs generally induce hypermutation and silencing of repetitive sequences through cytosine methylation[61]. In contrast, our analyses show that these outlier loci in *O. tauri* remain transcriptionally active and have low methylation, suggesting that alternative evolutionary and functional constraints, rather than centromere formation or a mechanism analogous to RIP-mediated silencing, underlie their distinct genomic architecture.

More broadly, the structural and epigenetic insulation of BOC1 and SOC1 underscores how such regions can function as supergenes where tightly linked genes accumulate alongside TEs and specialised regulatory mechanisms[62]. Because these outlier regions experience low recombination, the proliferation of TEs can be largely confined to these safe havens—minimising their impact on the standard genome and suggesting an intrinsic mechanism to harness, contain and domesticate TEs. In *Ostreococcus*, recombination suppression within the putative mating-type region preserves essential genes while tolerating elevated TE activity. Similarly, BOC1 and SOC1 may help Mamiellales maintain the energetic efficiency of a compact genome facing strong spatial constraints yet retain the flexibility to adapt in fluctuating marine environments[20].

In conclusion, our work shows that the outlier low-GC genomic regions are conserved across the Mamiellales lineage, where they form topologically insulated domains enriched in TEs, intron-rich genes, and with high expression levels. Our findings support the hypothesis that such outlier regions serve as “innovation hubs”: through enriched TE activity, alternative splicing, and epigenetic modifications, they expand metabolic and regulatory diversity without compromising the compactness of the standard genome. Although the full repertoire of functions encoded by these specialised genomic regions remains to be fully revealed, their enrichment in genes linked to secondary metabolism and antiviral defence strongly indicates a pivotal role in stress sensing and survival mechanisms. Indeed, the correlation between the SOC size and viral resistance in *Ostreococcus*[26]

parallels the potential function of BOC1 and SOC1 as reservoirs of adaptive genetic diversity. Our study points to a broader genomic mechanism, as outlier regions appear in Mamiellales and the distantly related stramenopile *Pelagomonas*. Taken together, our findings support a unifying model in which outlier genomic regions in diverse minimal eukaryotic genomes act as evolutionary test beds to harness TE-driven innovation and enhance metabolic flexibility for adaptation to dynamic marine ecosystems.

Methods

Algal strains and datasets

A total of eight Mamiellales species were included in the study, representing the genera *Ostreococcus*, *Bathycoccus*, *Micromonas* and *Mantoniella*. We used *Ostreococcus tauri* strain RCC4221 in this study to generate data for Hi-C, RNA-Seq and methylation profiling. We sourced other molecular datasets from publicly available databases (Supplementary Table 1).

Algal culturing

O. tauri strain RCC4221 was maintained in L1 seawater medium[63] in non-treated polycarbonate flasks and maintained in a 14 h: 10h light: dark cycle at 35 µmol m^{-2} s^{-1} irradiance and a temperature of 20 °C. To achieve a near-axenic culture of *O. tauri*, we isolated cells by flow cytometric cell sorting (BD Influx, Beckton Dickinson) and maintained algal cells in exponential growth by weekly subculturing. We also used a monthly antibiotic treatment (5 µg Ampicillin, 1 µg Gentamicin, 2 µg Kanamycin, 10 µg Neomycin). Algal growth was regularly monitored by flow cytometry (Cytoflex, Beckman Coulter) based on side-angle light scatter and chlorophyll red fluorescence. Bacterial cells were monitored and quantified via flow cytometry after SYBR Green I staining.

Transcriptomics

RNA extraction and sequencing. *O. tauri* RCC4221 cells were collected by centrifuging a 20 mL sample (10 min at 4000 × g at 4 °C), snap-frozen in liquid nitrogen, and stored at -80 °C. RNA was extracted using a RNeasy Plant kit (QIAGEN) with RLT buffer incubated at 56 °C for 3 min. The residual DNA was digested using the Turbo DNA-free kit (Thermo Fisher Scientific). RNA was quantified using a Qubit Fluorometer (Thermo Fisher Scientific), and RNA quality was assessed using an Agilent 2100 Bioanalyzer System. RNA extracts with an RNA integrity number (RIN) greater than eight were subjected to poly(A) selection prior to library preparation with the Illumina TruSeq Direction Library Prep Kit and sequenced on Illumina NovaSeq 6000 (150 bp paired end reads).

RNA-Seq data processing. Adapters and low-quality reads from RNA-Seq libraries were trimmed using Fastp v0.23.2[64] (minimum length required 50). Quality control reports were compiled using FastQC v0.11.9[65], MultiQC v1.13[66], and fastp. Ribosomal RNA reads were removed using SortMERNA v4.3.6[67] matching reads with rRNA

sequences from the SILVA databases[68]. The remaining reads were aligned and quantified to the *O. tauri* reference genome sequence (GCF_000214015.3) using Star v2.7.10[69] with default parameters. Counts were divided by gene length to standardise counts and generate transcript levels. For publicly available RNA-Seq libraries and counts, transcript levels were measured in *M. pusilla* CCMP1545 using table 'Sup_Table_S2_CCMP1545' column 'P-Replete T1.5 hours' for counts[70]. For *M. commoda* RCC299, the existing triplicate axenic counts[71]. For *B. prasinos* RCC1105 and *O. lucimarinus* CCE9901, the reads from SRR1300453[72] and SRR847308[38] were used with the same workflow as above.

Oxford Nanopore Technology (ONT) sequencing

DNA extraction and library preparation. Cells were collected by centrifuging a 30 mL sample (10 min at 4000 × g at 4 °C), snap-frozen in liquid nitrogen, and stored at -80 °C. DNA was extracted using a DNeasy Plant kit (QIAGEN). DNA was quantified using a Qubit Fluorometer (Thermo Fisher Scientific). Library preparation was conducted using the Ligation Sequencing Kits (SQK-LSK109; SQK-LSK114; ONT, Oxford, UK). Genomic DNA was pooled for sequencing on an FLO-FLG001 Flongle flow cell and FLO-MIN114 Minion flow cell (R9.4.1; R10.4.1; ONT, Oxford, UK).

Methylation profiling. Super-accuracy base calling was performed on R10.4.1 sequence data using Dorado v0.6.0 (ONT, Oxford, UK), Methylation profiling included N6-methyladenine (6mA), N5-methylcytosine (5mC/5mCG), and N5-hydroxymethylcytosine (5hmC/5hmCG) which were transformed into BED files using modkit v0.2.8 (ONT, Oxford, UK).

Genome assembly. Exeter RCC4221 was de novo assembled from R9.4.1 ONT sequence data using Flye v2.9.2[73] (--iterations 3 –scaffold) followed by medaka_consensus v1.8.0 (ONT, Oxford, UK). Matching scaffolds to reference were pulled out and aligned using blastn v2.1.0[74].

High-throughput chromosome conformation capture (Hi-C)

For consistency, samples for all Hi-C experiments were collected at 2 hours into the light phase. To ensure sufficient DNA yield for sequencing from small *O. tauri* genomes, 1–3 × 10^8 total cells were used in all experiments. As there was no published protocol for picoeukaryotes we tested both *in situ*[75] and *in vivo* Hi-C approaches and a series of optimisation steps, which are detailed in the Supplementary Materials. In summary, the final protocol was as follows: we performed crosslinking by adding 2% v/v final concentration formaldehyde (16% TEM grade, Sigma) directly within the culture medium containing the cells and incubating at room temperature for 20 mins. Cross-linking was stopped by adding glycine (0.2 M final concentration) and incubating for 5 mins at room temperature on a roller. Finally, cells were washed with ice-cold PBS and stored at -80 °C. The final Hi-C library was generated using the Proximo Microbe Hi-C kit v2 (PhaseGenomics, USA) following the

manufacturer's instructions and sequenced using the Illumina NovaSeq 6000 platform (150 bp paired end reads) yielding a total of 94143446 reads (run ID V0017_combined in Extended Data 3).

Hi-C data processing. Adapters and low-quality reads were trimmed from Hi-C libraries using Fastp v0.23.2. Quality control reports were compiled using FastQC v0.11.9[65], MultiQC v1.13[66], and fastp. The Hi-C processing pipeline, with the restriction sites of "^GATC,^AATT" and ligation sites of "GATCGATC,GATCAATT,AATTGATC,AATTAATT", was based on NextFlow v22.04.5 (available as a Docker profile: nf-core/hic v1.3.0)[76–78]. Hi-C reads were mapped onto a soft and hard-masked (Supplementary Fig. 18). *O. tauri* reference genome sequence (GenBank identifier: GCF_000214015.3) using Bowtie v2.3.5.1[79]. Duplication removal and contact pair validation were conducted with HiC-Pro v2.11.1[76]. Raw and normalised contact maps were generated using cooltools v0.3.0[80] at bin resolutions of 5, 10 and 20 kb. Additional normalisation was carried out using 'iced' with HiC-Pro, a Python implementation of the ICE normalisation algorithm from Cooltools[81]. Hi-C quality reports were generated using HiCUP[82]. Domain compartment calling was completed using Cooltools. The R package HiCRep[83] was used to evaluate the correlation between Hi-C experimental replicates (Supplementary Fig. 19). Juicer v1.6[84] was used to produce additional contact maps for use with downstream loop detection. Chromatin loops were identified from the Nextflow and Juicer contact maps using Mustache v1.0.1[85] and HiCCUPS v3.0[84,85], respectively, at bin resolutions of 5, 10 and 20kb., its derivative/slope and cis and trans interactions was calculated using cooltools. *M. tinhauana* RCC11003 Hi-C data[36] was similarly analysed[36].

MNase-seq analysis. MNase-seq data for *M. pusilla* CCMP1545 and *O. lucimarinus* CCE9901 (SRR847304 and SRR847307[38]) were processed to analyse nucleosome positioning. Paired end reads underwent quality trimming and adapter removal using fastp v0.23.4. The cleaned reads were aligned to the reference genome with Minimap2 v2.24[86] using default parameters for paired-end data. Aligned files were converted to BAM format, sorted, and indexed with Samtools v1.13[87]. Finally, genome-wide coverage was calculated and normalised to RPKM using bamCoverage (deeptools v3.5.1[88], --MNase --center-reads) to calculate nucleosome centers, and the results were exported as bigwig files for visualisation and downstream analysis in Python for peak calling. Autocorrelation functions of nucleosome center coverage were estimated using python. Nucleosome fuzziness and stiffness were calculated using Nucleosome Dynamics v1.0[89].

Bisulfite-seq analysis. Bisulfite sequencing data for *M. pusilla* CCMP1545 and *O. lucimarinus* CCE9901 (SRR847303 and SRR847306[38]) were processed to determine methylation levels at CpG sites. Single-end reads were trimmed for quality and adapters using Fastp v0.23.4. The trimmed reads were aligned to the reference genome using Bismark v0.24.2[90]. Methylation calling was restricted to CpG contexts, and methylation levels were summarised in bedgraph and bigwig format for further analysis.

Sequence data analyses. Protein synteny was highlighted using Oxford dot plots v0.3.3[91], using blastp v2.10[74] and Diamond v2.1.8[92] across outlier chromosomes in the Mamiellophyceae. A ribbon plot was generated to show the synteny between species using Python. EDTA v.2.2.2[93], default settings, were used to flag transposable elements across whole genomes and PKS genes, respectively. A data frame of all the proteins within *O. tauri* with GO enrichment spanning biological processes, molecular functions, cellular components, and locations on chromosomes was obtained from NCBI and UniProt[94]. Fisher's exact test (one-tailed) was performed using a custom script in Python on subsets of genes to compare enriched terms over target regions and the rest of the genome.

Metabolites. Metabolite profiling was carried out using LC-quadrupole time of flight mass spectrometry using separate runs for positive and negative ion modes. *O. tauri* RCC4221 and *O. lucimarinus* CCE9901 cells were collected by centrifuging a 30 mL sample (10 min at 4000 × g at 4 °C), snap-frozen in liquid nitrogen, and stored at -80 °C. Pellets were extracted in 0.6 ml methanol, treated for 20 min in a sonicating water bath with ice and then centrifuged at 20,000g for 10 min at 4 °C. The supernatants from three replicate cultures of each species and extraction blanks were analysed using an Agilent Zorbax Eclipse plus, narrow bore RR, 2.1 × 50 mm, 3.5 µm particle size reverse phase C18 analytical column (Agilent Technologies, Palo Alto, USA). 5 µL samples were injected *via* an Agilent 1200 series Rapid Resolution HPLC system coupled to a quadrupole time-of-flight (QToF) 6520 mass spectrometer (Agilent Technologies, Palo Alto, USA). Mobile phases were A (10% methanol, 0.1% formic acid), mobile phase B (100% methanol with 0.1% formic acid). The following gradient was used: 0–5 min-30% B; 10–24 min-0–98% B; 25-30 min-100% B. The flow rate was 0.3 mL min^{-1} and the column temperature was 40 °C. The source conditions for electrospray ionisation were as follows: gas temperature was 325 °C with a drying gas flow rate of 9 l min^{-1} and a nebuliser pressure of 35 psig. The capillary voltage was 3.5 kV in both positive and negative ion mode. The fragmentor voltage was 175 V and skimmer 65 V. Scanning was performed using the auto MS/MS function at 5 scans s^{-1} for precursor ion surveying and 4 scans s^{-1} for MS/MS with a sloped collision energy of 3.5 V/100 Da with an offset of 10 V. MS and MS/MS m/z range was 40-1700. Reference compound m/z values were 121.05087 and 922.00979. Raw data files (*.d) were converted to mzML format using MSConvert (ProteoWizard)[95]. Feature detection and alignment was carried out with MS-Dial v5[96]. Features present in the blank (>20% of sample intensity) were discounted. Compounds were identified with the MS-Dial lipidomics workflow and using publicly available CompMS MS-Dial MS/MS libraries. Further runs were carried out to obtain additional MS/MS data to verify potential polyketide products. Metabolite products were predicted from protein sequences (XP_001416378.1 and XP_022840687.1) using ClusterCAD v2.0[97] and SeMPI v2.0[98] and PKS domains predicted using NCBI Conserved Domain Search[99] and antiSMASH[100]. Potential adduct types were calculated from the exact masses of these putative products taking entries with m/z withing 0.005 of predicted exact masses. Enrichment tests were done using Kruskal Wallis.

Supplementary Information

In supplementary file.

Data availability

Zenodo Link : 10.5281/zenodo.14284467

NCBI Link: PRJNA1250464

Acknowledgments

Illumina sequencing was conducted on equipment funded by the Wellcome Trust (Multi-User Equipment Grant award number 218247/Z/19/Z). We are grateful to Dr Karen Moore (Exeter Sequencing Facility, University of Exeter) for their technical advice on Hi-C sequencing. We thank Aurélie Chambouvet and Evelyne Derelle for providing algal strains. This research was supported by the Royal Society through a University Research Fellowship (URF\R1\180537) and Research Fellows Enhancement Awards (RGF\EA\181012) awarded to AM.

Figure legends

Figure 1. Comparative genomic analysis of BOC1 regions across Mamiellales species.

(**a**), Syntenic analysis and gene homology visualisation of BOC1 regions across six Mamiellales species. Connecting lines indicate homologous genes identified through reciprocal best BLASTP hits between *O. tauri* RCC4221, *O.* sp. RCC809, *O. lucimarinus* CCE9901, *B. prasinos* RCC1105, *M. pusilla* CCMP1545, and *M. commoda* RCC299. BOC1 regions are displayed in green. BOC1-localised orthologues are indicated by green lines. Total BOC1 length and the proportion contained within the region (in parentheses) are indicated for each species. (**b**), Gene Ontology (GO) enrichment analysis of BOC1-encoded proteins in RCC4221. The plot displays GO terms categorised into three major classes: biological process (pink), cellular component (blue), and molecular function (yellow). Enrichment values are shown as $-\text{Log}_{10}$(p-value), with a significance threshold of $p<0.0335$ (Fisher's exact test). Dot size indicates gene count, while colour intensity represents the odds ratio, defined as the proportion of GO term occurrence in BOC1 proteins relative to non-BOC1 regions.

Figure 2. Distribution and expression of transposable elements across Mamiellales genomes.

(**a**), Dendrogram of major transposable element (TE) classes in Mamiellales. The left sidebars indicate the classification of elements into retrotransposon, with further subdivision into LTR/non-LTR retrotransposons and DDE/D transposons, rolling-circle elements, and YR cryptons. Heatmap shows TE density (TEs MB^{-1}) across standard (S), BOC1 (B), and SOC1 (S) genomic regions. The lower dendrogram displays phylogenetic relationships between Mamiellales species: *O. tauri* (Ot; strains RCC4221 and RCC1115), *O.* sp. RCC809 (Os), *O. lucimarinus* (Ol; CCE9901), *B. prasinos* (Bp; RCC1105), *M. pusilla* (Mp; CCMP1545), *M. commoda* (Mc;

RCC299), and *M. tinhauana* (Mt; RCC11003). (**b**), Quantitative comparison of TE abundance (TEs MB^{-1}) across different genomic regions. Bar colours correspond to standard (grey), BOC1 (green), and SOC1 (red) regions. (**c—f**), RNA-seq expression profiles of representative TEs in *O. tauri* RCC4221 showing: (c) Ty1/Copia element localised upstream of a putative hexosyltransferase gene; (d) Tc1/Mariner element near a transcription factor IIA; (e and f) Mutator elements near polyketide synthases. Grey peaks represent RNA-Seq coverage, with maximum coverage values indicated in the top right corner of each track. Dotted lines represent 200x coverage threshold. Grey bars indicate predicted protein-coding regions with their corresponding NCBI locus tags (genome assembly GCF_000214015.3). Updated protein annotations were determined using AlphaFold3[101] and FoldSeek[102] analysis.

Figure 3. Higher gene expression in BOC1.

(**a—d**), Expression levels, GC content and TE content across BOC of (**a**) *O. tauri* (RCC4221; chromosome 2), (**b**), *M. commoda* (RCC299; chromosome 1), and (**c**) *B. prasinos* (RCC1105; chromosome 14). Boundary points of the BOC1 regions are labelled with cyan dotted lines; transposable elements are shown on the lower track. (**d—f**), Comparative analysis of gene expression in outlier regions and standard genome; box plots display the distribution of transcript levels across outlier (BOC1 and SOC1) and standard regions for each taxon (Kruskal-Wallis, $p<0.001$). (**g—i**), Correlation between gene expression and GC content of coding sequences (CDS) in (**g**), *O. tauri* (Spearman's rank, $\rho=-0.366$, $p=7.82e^{-242}$), (**h**), *M. commoda* ($\rho = -0.285$, $p = 3.70e^{-192}$), and (i) *B. prasinos* ($\rho = -0.38302$, $p = 8.49e^{-13}$); CDS in BOC1 and non-BOC1 regions are represented in green and blue, respectively.

Figure 4. Chromosomal organisation and three-dimensional genome architecture in *O. tauri*.

(**a**), Hi-C contact map of the *O. tauri* genome binned at 10 kb resolution with normalised interactions counts; chromosomes 2 and 19, which display enriched intrachromosomal contacts, highlighted by white squares.
(**b**), Hi-C normalised contact map of chromosome 2 at 20 kb resolution. The first track displays transcript levels alongside GC content. The second track shows the rolling average of DNA modifications within a scaled window, 5mCG (orange), 5mC (green), 6mA (purple), 5hmCG (blue). The third track indicates positions of transposable elements (black).
(**c—d**), Hi-C contact maps binned at 10 kb for (**c**) chromosome 2 and (**d**) chromosome 19. Topologically associating domains are indicated by black dotted lines, with outlier regions labelled as BOC1 and SOC1 (cyan text), the bottom track represents the insulation score calculated as a rolling z-score from the summation of Hi-C contacts at 5 kb (red), 10 kb (blue), and 20 kb (green) resolutions. (**e**), Distribution of normalised interchromosomal contact probabilities between different chromosomes. Interactions involving BOC or SOC chromosomes are labelled in green, whilst interactions between standard nuclear are shown in blue. Notable interactions between chromosomes 2 and 19, and between chromosomes 18 and 20 are highlighted.

Figure 5. **MNase and CpG profiles across BOC1 in *Ostreococcus lucimarinus* CCE9901 and *Micromonas pusilla* CCMP1545.**

(**a, b**), Transcript levels, CpG methylation as a percentage of base coverage which is methylated, MNase coverage alongside TEs and IEs in *O. lucimarinus* CCE9901 (**a**) and *M. pusilla* CCMP1545 (**b**). Dotted lines indicate the edge of outlier regions. Blue bars indicate PKS coding regions. (**c, e**), Means at each position aligned to transcription start sites are shown for CpG methylation and nucleosomes within the standard genome above and BOC1 genes below for *O. lucimarinus* (**c**) and *M. pusilla* (**e**). Blue is the percentage of bases which are CpG methylated. Black is nucleosome center counts. Light grey is MNase coverage. (**d, f**), The autocorrelation function estimates of nucleosome center counts are shown for each lag (offset, bp) across standard genome (black) and BOC1 (turquoise) genes for *O. lucimarinus* (**d**) and *M. pusilla* (**f**), respectively. Period is annotated in base pairs.

Figure 6. **Modular Giant Type 1 Polyketide synthase present in Mamiellales.**

(**a**), Schematic representation of the largest orthologous polyketide synthases (PKSs) identified across Mamiellales species. (**b**), Hierarchical clustered heatmap showing log2 abundances of putative polyketide products (metabolites) across triplicate cultures of *O. tauri* and *O. lucimarinus*. Blue and red rectangles highlight products significantly enriched in *O. tauri* and *O. lucimarinus*, respectively (Kruskal-Wallis, $p<0.05$). (**c**), Molecular diagrams of selected enriched polyketides products from *O. tauri* and *O. lucimarinus*. Mycolactone, Ebelactone A, 6-Deoxyerythronolide B, Actinoallolide A.

References

1. Courties, C. *et al.* Smallest eukaryotic organism [6]. *Nature* vol. 370 Preprint at https://doi.org/10.1038/370255a0 (1994).
2. Chretiennot-Dinet, M. J. *et al.* A new marine picoeucaryote: Ostreococcus tauri gen. et sp. nov. (Chlorophyta, Prasinophyceae). *Phycologia* **34**, (1995).
3. Noordally, Z. B. *et al.* A phospho-dawn of protein modification anticipates light onset in the picoeukaryote Ostreococcus tauri. *J Exp Bot* **74**, (2023).
4. Corellou, F. *et al.* Clocks in the green lineage: Comparative functional analysis of the circadian architecture of the picoeukaryote ostreococcus. *Plant Cell* **21**, (2009).
5. Martin, S. F. *et al.* Surviving starvation: Proteomic and lipidomic profiling of nutrient deprivation in the smallest known free-living eukaryote. *Metabolites* **10**, (2020).
6. Marin, B. & Melkonian, M. Molecular Phylogeny and Classification of the Mamiellophyceae class. nov. (Chlorophyta) based on Sequence Comparisons of the Nuclear- and Plastid-encoded rRNA Operons. *Protist* **161**, (2010).

7. Tragin, M. & Vaulot, D. Novel diversity within marine Mamiellophyceae (Chlorophyta) unveiled by metabarcoding. *Sci Rep* **9**, (2019).
8. Yung, C. C. M., Redondo, E. R., Sanchez, F., Yau, S. & Piganeau, G. Diversity and Evolution of Mamiellophyceae: Early-Diverging Phytoplanktonic Green Algae Containing Many Cosmopolitan Species. *Journal of Marine Science and Engineering* vol. 10 Preprint at https://doi.org/10.3390/jmse10020240 (2022).
9. Demory, D. *et al.* Picoeukaryotes of the Micromonas genus: sentinels of a warming ocean. *ISME Journal* **13**, (2019).
10. Demir-Hilton, E. *et al.* Global distribution patterns of distinct clades of the photosynthetic picoeukaryote Ostreococcus. *ISME Journal* **5**, (2011).
11. Monier, A., Worden, A. Z. & Richards, T. A. Phylogenetic diversity and biogeography of the Mamiellophyceae lineage of eukaryotic phytoplankton across the oceans. *Environ Microbiol Rep* **8**, (2016).
12. Sands, E. *et al.* Genetic and physiological responses to light quality in a deep ocean ecotype of Ostreococcus, an ecologically important photosynthetic picoeukaryote. *J Exp Bot* **74**, (2023).
13. Merchant, S. S. *et al.* The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science (1979)* **318**, (2007).
14. Derelle, E. *et al.* Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features. *Proc Natl Acad Sci U S A* **103**, (2006).
15. Palenik, B. *et al.* The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. *Proc Natl Acad Sci U S A* **104**, (2007).
16. Worden, A. Z. *et al.* Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes micromonas. *Science (1979)* **324**, (2009).
17. van Baren, M. J. *et al.* Evidence-based green algal genomics reveals marine diversity and ancestral characteristics of land plants. *BMC Genomics* **17**, (2016).
18. Devic, M. *et al.* An INDEL genomic approach to explore population diversity of phytoplankton. *BMC Genomics* **25**, 1045 (2024).
19. Da Silva, O. *et al.* Genomic differentiation of three pico-phytoplankton species in the Mediterranean Sea. *Environ Microbiol* **24**, (2022).
20. Giovannoni, S. J., Cameron Thrash, J. & Temperton, B. Implications of streamlining theory for microbial ecology. *ISME Journal* vol. 8 Preprint at https://doi.org/10.1038/ismej.2014.60 (2014).
21. Blanc-Mathieu, R. *et al.* An improved genome of the model marine alga Ostreococcus tauri unfolds by assessing Illumina de novo assemblies. *BMC Genomics* **15**, (2014).

22. Grimsley, N., Yau, S., Piganeau, G. & Moreau, H. Typical features of genomes in the mamiellophyceae. in *Marine Protists: Diversity and Dynamics* (2015). doi:10.1007/978-4-431-55130-0_6.

23. Moreau, H. *et al*. Gene functionalities and genome structure in Bathycoccus prasinos reflect cellular specializations at the base of the green lineage. *Genome Biol* **13**, (2012).

24. Lanier, W., Moustafa, A., Bhattacharya, D. & Comeron, J. M. EST analysis of Ostreococcus lucimarinus, the most compact eukaryotic genome, shows an excess of introns in highly expressed genes. *PLoS One* **3**, (2008).

25. Irimia, M. & Roy, S. W. Evolutionary convergence on highly-conserved 3' intron structures in intron-poor eukaryotes and insights into the ancestral eukaryotic genome. *PLoS Genet* **4**, (2008).

26. Blanc-Mathieu, R. *et al*. Population genomics of picophytoplankton unveils novel chromosome hypervariability. *Sci Adv* **3**, (2017).

27. Yau, S. *et al*. A Viral Immunity Chromosome in the Marine Picoeukaryote, Ostreococcus tauri. *PLoS Pathog* **12**, (2016).

28. Benites, L. F. *et al*. Evolutionary Genomics of Sex-Related Chromosomes at the Base of the Green Lineage. *Genome Biol Evol* **13**, (2021).

29. Yau, S. *et al*. Virus-host coexistence in phytoplankton through the genomic lens. *Sci Adv* **6**, (2020).

30. Lawson, H. A., Liang, Y. & Wang, T. Transposable elements in mammalian chromatin organization. *Nature Reviews Genetics* vol. 24 Preprint at https://doi.org/10.1038/s41576-023-00609-6 (2023).

31. Zhang, A. & Zhang, W. Characterization of Transposon-Derived Accessible Chromatin Regions in Rice (Oryza Sativa). *Int J Mol Sci* **23**, (2022).

32. Dekker, J. GC- and AT-rich chromatin domains differ in conformation and histone modification status and are differentially modulated by Rpd3p. *Genome Biol* **8**, (2007).

33. Vinogradov, A. E. DNA helix: The importance of being GC-rich. *Nucleic Acids Research* vol. 31 Preprint at https://doi.org/10.1093/nar/gkg296 (2003).

34. Barahimipour, R. *et al*. Dissecting the contributions of GC content and codon usage to gene expression in the model alga Chlamydomonas reinhardtii. *Plant Journal* **84**, (2015).

35. Niu, D. K. & Yang, Y. F. Why eukaryotic cells use introns to enhance gene expression: Splicing reduces transcription-associated mutagenesis by inhibiting topoisomerase I cutting activity. *Biol Direct* **6**, (2011).

36. Rey Redondo, E., Xu, Y. & Yung, C. C. M. Genomic characterisation and ecological distribution of Mantoniella tinhauana: a novel Mamiellophycean green alga from the Western Pacific. *Front Microbiol* **15**, 1358574 (2024).

37. Jancek, S., Gourbière, S., Moreau, H. & Piganeau, G. Clues about the genetic basis of adaptation emerge from comparing the proteomes of two Ostreococcus ecotypes (Chlorophyta, Prasinophyceae). *Mol Biol Evol* **25**, (2008).

38. Huff, J. T. & Zilberman, D. Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. *Cell* **156**, (2014).
39. Verhelst, B., Van De Peer, Y. & Rouzé, P. The complex intron landscape and massive intron invasion in a picoeukaryote provides insights into intron evolution. *Genome Biol Evol* **5**, (2013).
40. Shelest, E., Heimerl, N., Fichtner, M. & Sasso, S. Multimodular type I polyketide synthases in algae evolve by module duplications and displacement of AT domains in trans. *BMC Genomics* **16**, 1–15 (2015).
41. Guérin, N. *et al.* Genomic adaptation of the picoeukaryote Pelagomonas calceolata to iron-poor oceans revealed by a chromosome-scale genome sequence. *Commun Biol* **5**, (2022).
42. Guérin, N. *et al.* Transcriptomic response of the picoalga Pelagomonas calceolata to nitrogen availability: new insights into cyanate lyase function. *Microbiol Spectr* (2025) doi:10.1128/SPECTRUM.02654-24.
43. Henderson, G. P., Gan, L. & Jensen, G. J. 3-D Ultrastructure of O. tauri: Electron Cryotomography of an Entire Eukaryotic Cell. *PLoS One* **2**, e749 (2007).
44. Das, S. & Bansal, M. Variation of gene expression in plants is influenced by gene architecture and structural properties of promoters. *PLoS One* **14**, (2019).
45. Blattler, A. & Farnham, P. J. Cross-talk between site-specific transcription factors and DNA methylation states. *Journal of Biological Chemistry* vol. 288 Preprint at https://doi.org/10.1074/jbc.R113.512517 (2013).
46. Kribelbauer, J. F., Lu, X. J., Rohs, R., Mann, R. S. & Bussemaker, H. J. Toward a Mechanistic Understanding of DNA Methylation Readout by Transcription Factors. *Journal of Molecular Biology* vol. 432 Preprint at https://doi.org/10.1016/j.jmb.2019.10.021 (2020).
47. Luo, C., Hajkova, P. & Ecker, J. R. Dynamic DNA methylation: In the right place at the right time. *Science* vol. 361 Preprint at https://doi.org/10.1126/science.aat6806 (2018).
48. Ohtani, H. & Iwasaki, Y. W. Rewiring of chromatin state and gene expression by transposable elements. *Development Growth and Differentiation* vol. 63 Preprint at https://doi.org/10.1111/dgd.12735 (2021).
49. Huda, A., Mariño-Ramírez, L., Landsman, D. & Jordan, I. K. Repetitive DNA elements, nucleosome binding and human gene expression. *Gene* **436**, (2009).
50. Quadrana, L. *et al.* Transposition favors the generation of large effect mutations that may facilitate rapid adaption. *Nat Commun* **10**, (2019).
51. Arkhipova, I. R. Neutral theory, transposable elements, and eukaryotic genome evolution. *Mol Biol Evol* **35**, (2018).
52. Dupeyron, M., Singh, K. S., Bass, C. & Hayward, A. Evolution of Mutator transposable elements across eukaryotic diversity. *Mob DNA* **10**, (2019).
53. Nivina, A., Yuet, K. P., Hsu, J. & Khosla, C. Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chemical Reviews* vol. 119 Preprint at https://doi.org/10.1021/acs.chemrev.9b00525 (2019).

54. Zeng, G. *et al.* Duplication of a Pks gene cluster and subsequent functional diversification facilitate environmental adaptation in Metarhizium species. *PLoS Genet* **14**, (2018).

55. Heimerl, N. *et al.* A giant type I polyketide synthase participates in zygospore maturation in Chlamydomonas reinhardtii. *Plant J* **95**, 268–281 (2018).

56. Fallon, T. R. *et al.* Giant polyketide synthase enzymes in the biosynthesis of giant marine polyether toxins. *Science* **385**, 671–678 (2024).

57. Grimsley, N., Péquin, B., Bachy, C., Moreau, H. & Piganeau, G. Cryptic sex in the smallest eukaryotic marine green alga. *Mol Biol Evol* **27**, (2010).

58. Charlesworth, D. Evolution of recombination rates between sex chromosomes. *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 372 Preprint at https://doi.org/10.1098/rstb.2016.0456 (2017).

59. Gozashti, L. *et al.* Transposable elements drive intron gain in diverse eukaryotes. *Proc Natl Acad Sci U S A* **119**, (2022).

60. Bloom, K. S. Centromeric heterochromatin: The primordial segregation machine. *Annu Rev Genet* **48**, (2014).

61. Irelan, J. T. & Selker, E. U. Cytosine methylation associated with repeat-induced point mutation causes epigenetic gene silencing in Neurospora crassa. *Genetics* **146**, (1997).

62. Berdan, E. L. *et al.* Genomic architecture of supergenes: Connecting form and function. *Philosophical Transactions of the Royal Society B: Biological Sciences* **377**, (2022).

63. Guillard, R. R. L. & Hargraves, P. E. Stichochrysis immobilis is a diatom, not a chrysophyte. *Phycologia* **32**, (1993).

64. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. in *Bioinformatics* vol. 34 (2018).

65. Andrews, S. FastQC. *Babraham Bioinformatics* (2010).

66. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, (2016).

67. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, (2012).

68. Glöckner, F. O. The SILVA Database Project: An ELIXIR core data resource for high-quality ribosomal RNA sequences. *Biodiversity Information Science and Standards* **3**, (2019).

69. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, (2013).

70. Bachy, C. *et al.* Transcriptional responses of the marine green alga Micromonas pusilla and an infecting prasinovirus under different phosphate conditions. *Environ Microbiol* **20**, 2898–2912 (2018).

71. Ferrer-González, F. X. *et al.* Bacterial transcriptional response to labile exometabolites from photosynthetic picoeukaryote Micromonas commoda. *ISME Communications* **3**, (2023).
72. Keeling, P. J. *et al.* The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLoS Biol* **12**, (2014).
73. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* **37**, (2019).
74. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, (2009).
75. Belton, J. M. *et al.* Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* **58**, (2012).
76. Servant, N. *et al.* HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, (2015).
77. DI Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nature Biotechnology* vol. 35 Preprint at https://doi.org/10.1038/nbt.3820 (2017).
78. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology* vol. 38 Preprint at https://doi.org/10.1038/s41587-020-0439-x (2020).
79. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, (2012).
80. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).
81. Abdennur, N. *et al.* Cooltools: Enabling high-resolution Hi-C analysis in Python. *PLoS Comput Biol* **20**, (2024).
82. Wingett, S. *et al.* HiCUP: Pipeline for mapping and processing Hi-C data. *F1000Res* **4**, (2015).
83. Yang, T. *et al.* HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res* **27**, (2017).
84. Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, (2016).
85. Roayaei Ardakany, A., Gezer, H. T., Lonardi, S. & Ay, F. Mustache: Multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. *Genome Biol* **21**, (2020).
86. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, (2018).
87. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, (2009).
88. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. DeepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, (2014).
89. Buitrago, D. *et al.* Nucleosome Dynamics: a new tool for the dynamic analysis of nucleosome positioning. *Nucleic Acids Res* **47**, (2019).

90. Krueger, F. & Andrews, S. R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, (2011).

91. Schultz, D. T. *et al.* Ancient gene linkages support ctenophores as sister to other animals. *Nature* **618**, (2023).

92. Buchfink, B., Reuter, K. & Drost, H. G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* **18**, (2021).

93. Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol* **20**, (2019).

94. MacDougall, A. *et al.* UniRule: A unified rule resource for automatic annotation in the UniProt knowledgebase. *Bioinformatics* **36**, (2020).

95. Chambers, M. C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology* vol. 30 Preprint at https://doi.org/10.1038/nbt.2377 (2012).

96. Tsugawa, H. *et al.* MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods* **12**, (2015).

97. Eng, C. H. *et al.* ClusterCAD: A computational platform for type i modular polyketide synthase design. *Nucleic Acids Res* **46**, (2018).

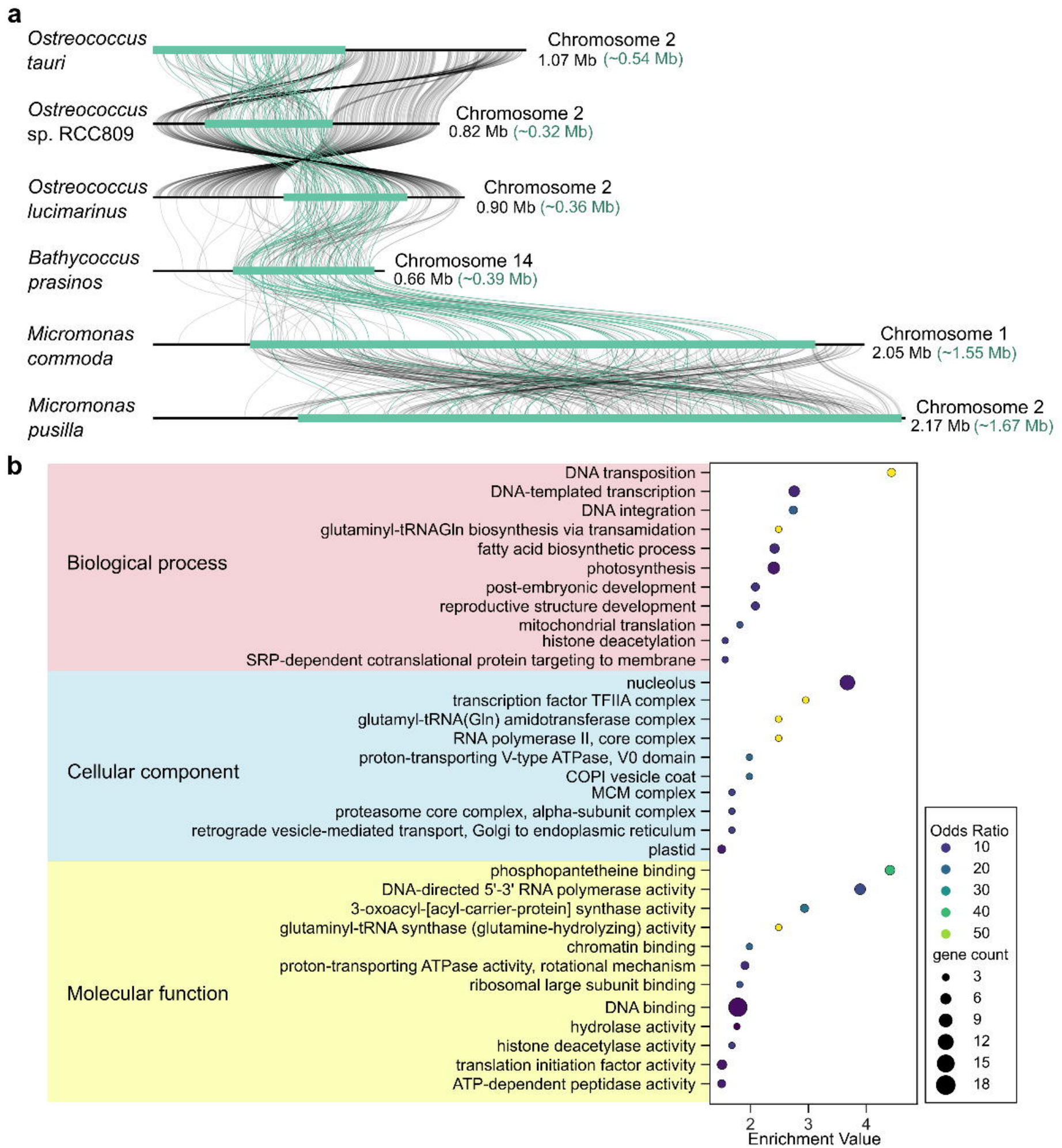
98. Zierep, P. F. *et al.* SeMPI: A genome-based secondary metabolite prediction and identification web server. *Nucleic Acids Res* **45**, (2017).

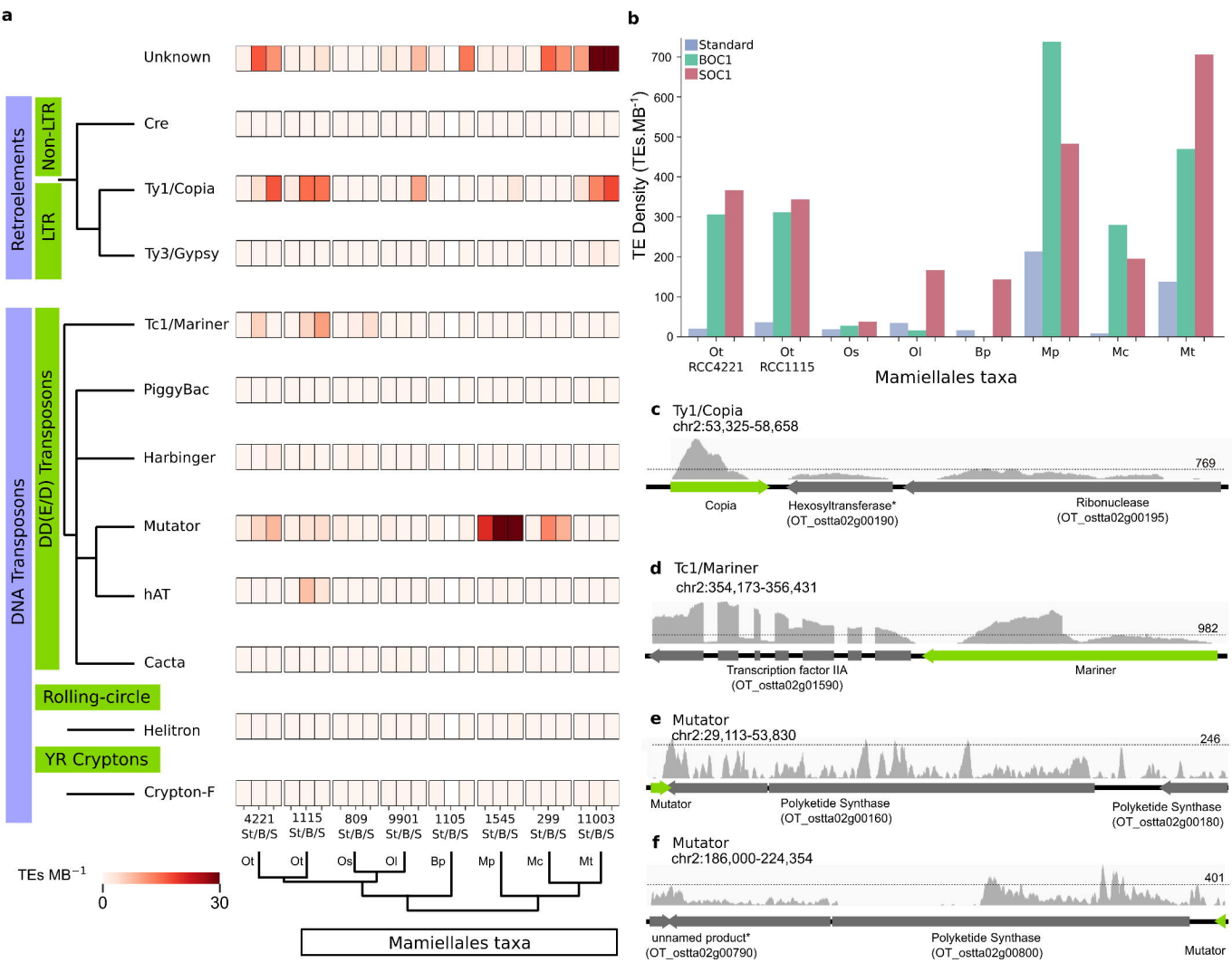
99. Wang, J. *et al.* The conserved domain database in 2023. *Nucleic Acids Res* **51**, (2023).

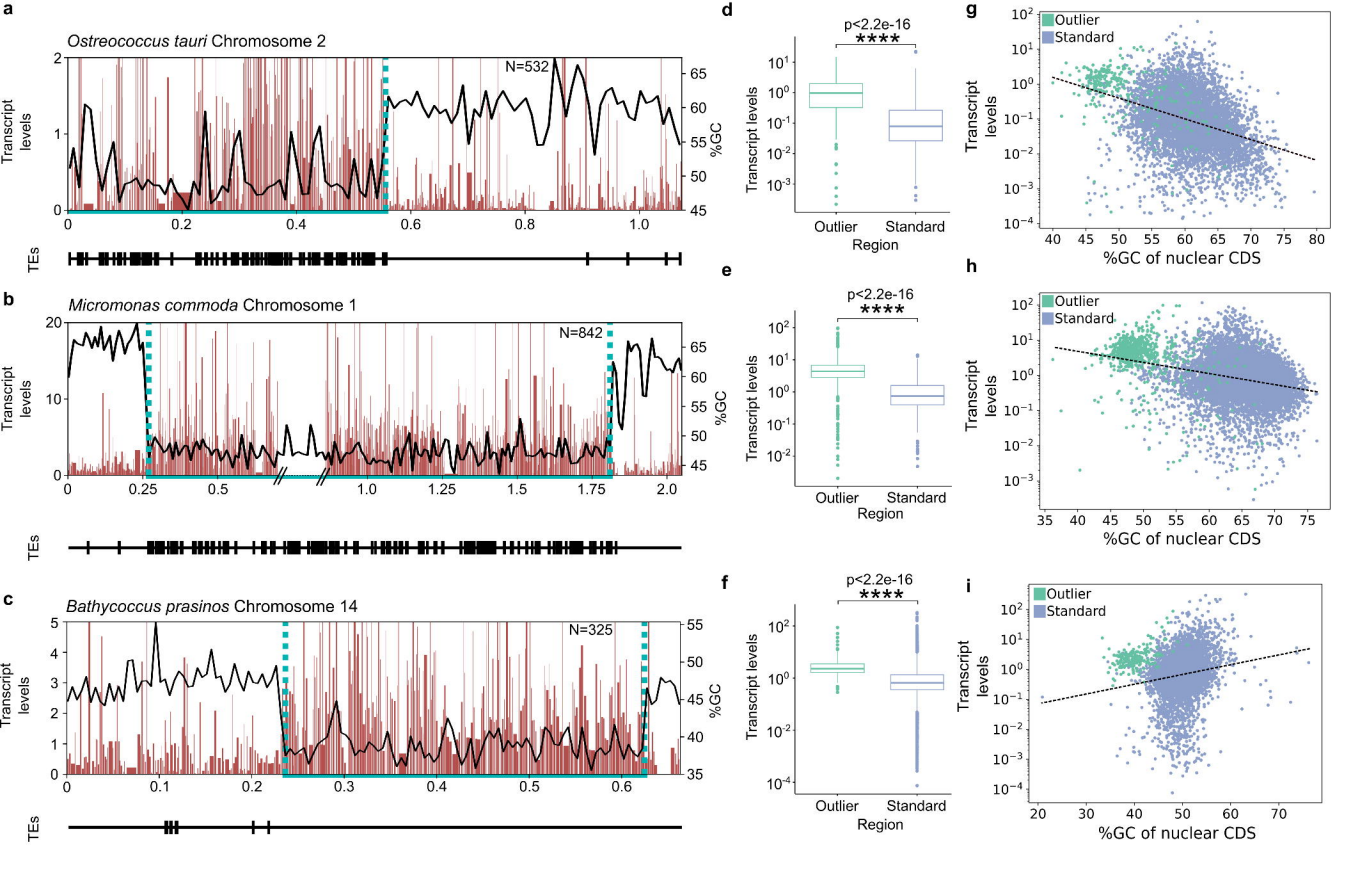
100. Blin, K. *et al.* AntiSMASH 7.0: New and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res* **51**, (2023).

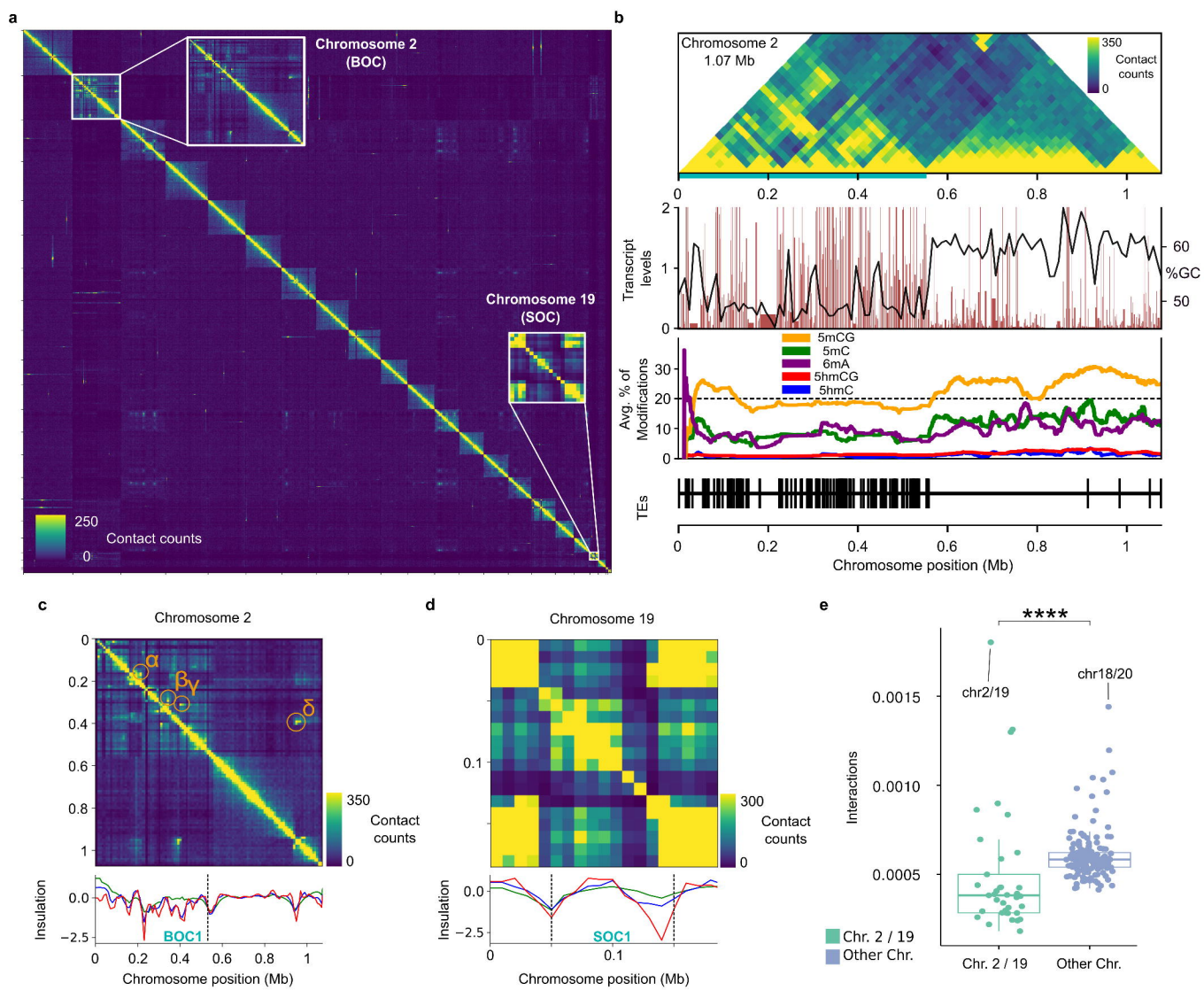
101. Varadi, M. *et al.* AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50**, (2022).

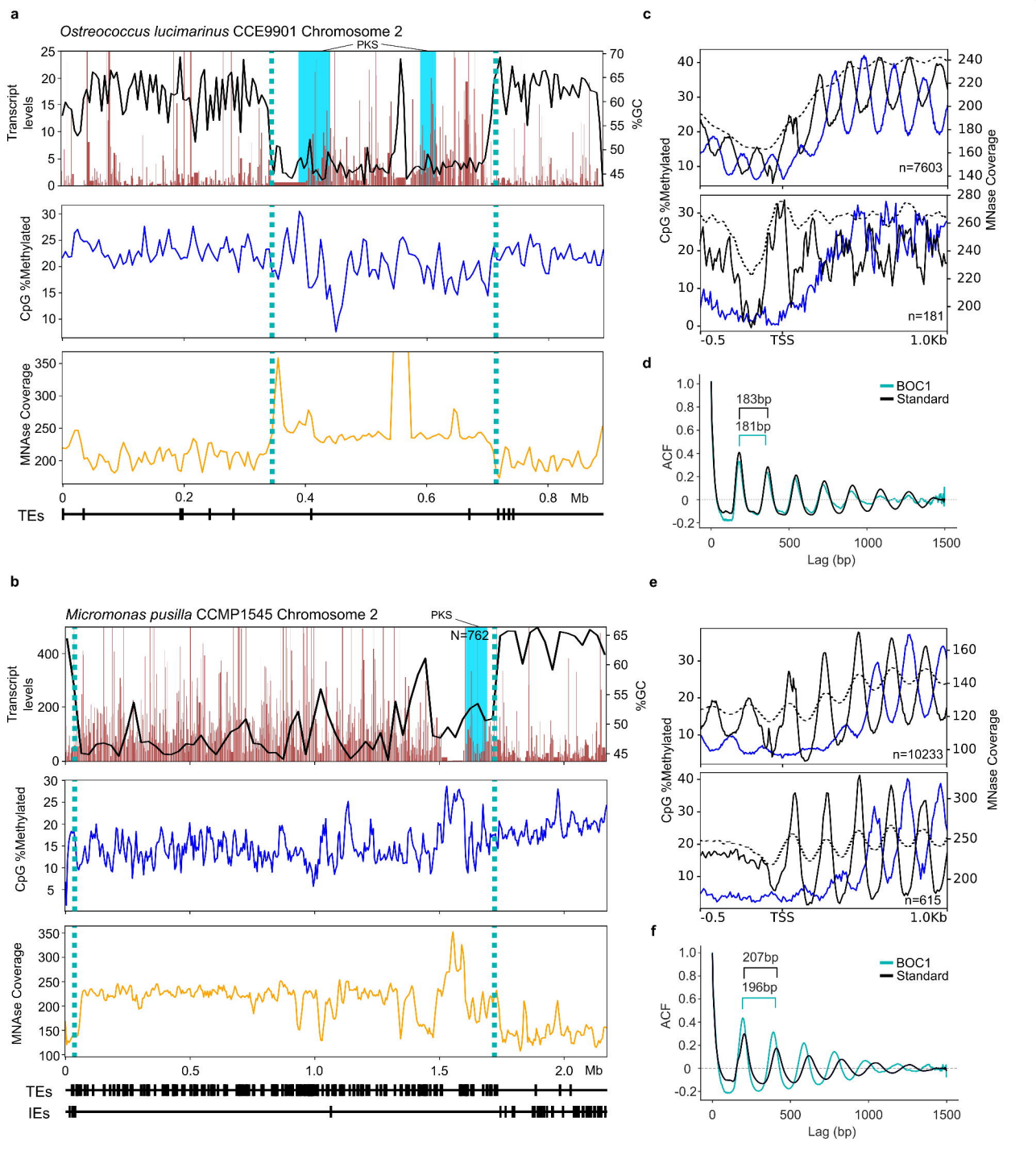
102. van Kempen, M. *et al.* Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* **42**, (2024).

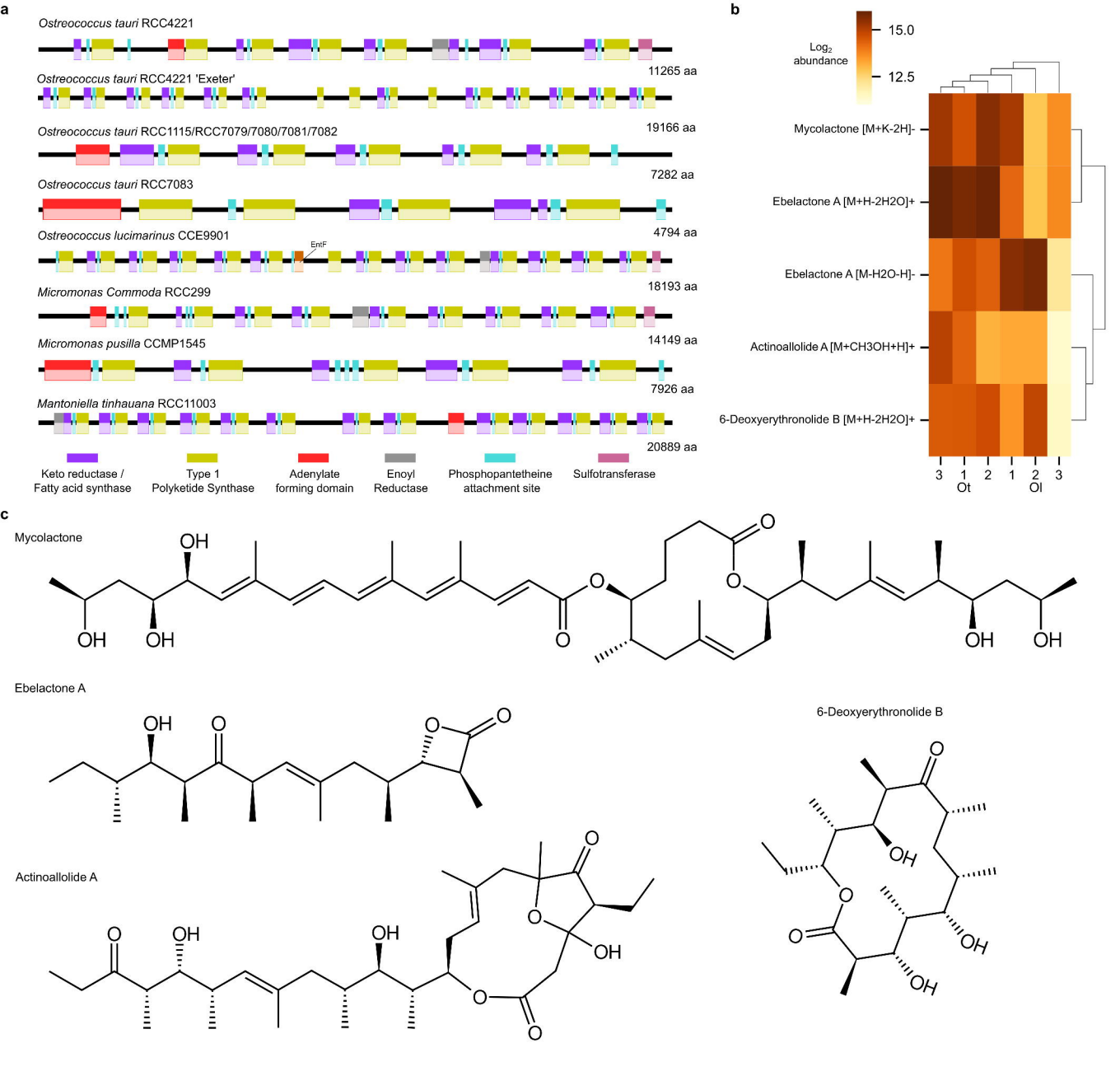

a
Ostreococcus tauri
Chromosome 2
1.07 Mb (~0.54 Mb)
Ostreococcus sp. RCC809
Chromosome 2
0.82 Mb (~0.32 Mb)
Ostreococcus lucimarinus
Chromosome 2
0.90 Mb (~0.36 Mb)
Bathycoccus prasinos
Chromosome 14
0.66 Mb (~0.39 Mb)
Micromonas commoda
Chromosome 1
2.05 Mb (~1.55 Mb)
Micromonas pusilla
Chromosome 2
2.17 Mb (~1.67 Mb)
b
Biological process
DNA transposition
DNA-templated transcription
DNA integration
glutaminyl-tRNAGln biosynthesis via transamidation
fatty acid biosynthetic process
photosynthesis
post-embryonic development
reproductive structure development
mitochondrial translation
histone deacetylation
SRP-dependent cotranslational protein targeting to membrane
Cellular component
nucleolus
transcription factor TFIIA complex
glutamyl-tRNA(Gln) amidotransferase complex
RNA polymerase II, core complex
proton-transporting V-type ATPase, V0 domain
COPI vesicle coat
MCM complex
proteasome core complex, alpha-subunit complex
retrograde vesicle-mediated transport, Golgi to endoplasmic reticulum
plastid
Molecular function
phosphopantetheine binding
DNA-directed 5'-3' RNA polymerase activity
3-oxoacyl-[acyl-carrier-protein] synthase activity
glutaminyl-tRNA synthase (glutamine-hydrolyzing) activity
chromatin binding
proton-transporting ATPase activity, rotational mechanism
ribosomal large subunit binding
DNA binding
hydrolase activity
histone deacetylase activity
translation initiation factor activity
ATP-dependent peptidase activity
2
3
4
Enrichment Value
Odds Ratio
10
20
30
40
50
gene count
3
6
9
12
15
18


a
Unknown
Retroelements
Non-LTR
Cre
LTR
Ty1/Copia
Ty3/Gypsy
DNA Transposons
DD(E/D) Transposons
Tc1/Mariner
PiggyBac
Harbinger
Mutator
hAT
Cacta
Rolling-circle
Helitron
YR Cryptons
Crypton-F
4221 St/B/S
1115 St/B/S
809 St/B/S
9901 St/B/S
1105 St/B/S
1545 St/B/S
299 St/B/S
11003 St/B/S
Ot
Ot
Os
Ol
Bp
Mp
Mc
Mt
TEs MB^{-1}
0
30
Mamiellales taxa
b
Standard
BOC1
SOC1
TE Density (TEs.MB^{-1})
0
100
200
300
400
500
600
700
Ot RCC4221
Ot RCC1115
Os
Ol
Bp
Mp
Mc
Mt
Mamiellales taxa
c Ty1/Copia
chr2:53,325-58,658
769
Copia
Hexosyltransferase*
(OT_ostta02g00190)
Ribonuclease
(OT_ostta02g00195)
d Tc1/Mariner
chr2:354,173-356,431
982
Transcription factor IIA
(OT_ostta02g01590)
Mariner
e Mutator
chr2:29,113-53,830
246
Mutator
Polyketide Synthase
(OT_ostta02g00160)
Polyketide Synthase
(OT_ostta02g00180)
f Mutator
chr2:186,000-224,354
401
unnamed product*
(OT_ostta02g00790)
Polyketide Synthase
(OT_ostta02g00800)
Mutator


a
Ostreococcus tauri Chromosome 2
N=532
Transcript levels
%GC
TEs
b
Micromonas commoda Chromosome 1
N=842
Transcript levels
%GC
TEs
c
Bathycoccus prasinos Chromosome 14
N=325
Transcript levels
%GC
TEs
d
p<2.2e-16
****
Transcript levels
Outlier
Standard
Region
e
p<2.2e-16
****
Transcript levels
Outlier
Standard
Region
f
p<2.2e-16
****
Transcript levels
Outlier
Standard
Region
g
Outlier
Standard
Transcript levels
%GC of nuclear CDS
h
Outlier
Standard
Transcript levels
%GC of nuclear CDS
i
Outlier
Standard
Transcript levels
%GC of nuclear CDS


a
Chromosome 2
(BOC)
Chromosome 19
(SOC)
250
Contact counts
0
b
Chromosome 2
1.07 Mb
350
Contact
counts
0
Transcript
levels
%GC
5mCG
5mC
6mA
5hmCG
5hmC
Avg. % of
Modifications
TEs
Chromosome position (Mb)
c
Chromosome 2
α
β
γ
δ
350
Contact
counts
0
Insulation
BOC1
Chromosome position (Mb)
d
Chromosome 19
300
Contact
counts
0
Insulation
SOC1
Chromosome position (Mb)
e
****
chr2/19
chr18/20
Interactions
Chr. 2 / 19
Other Chr.


a
Ostreococcus lucimarinus CCE9901 Chromosome 2
PKS
Transcript levels
%GC
CpG %Methylated
MNAse Coverage
TEs
Mb
b
Micromonas pusilla CCMP1545 Chromosome 2
PKS
N=762
Transcript levels
%GC
CpG %Methylated
MNAse Coverage
TEs
IEs
Mb
c
CpG %Methylated
MNase Coverage
n=7603
n=181
-0.5
TSS
1.0Kb
d
BOC1
Standard
183bp
181bp
ACF
Lag (bp)
e
CpG %Methylated
MNase Coverage
n=10233
n=615
-0.5
TSS
1.0Kb
f
BOC1
Standard
207bp
196bp
ACF
Lag (bp)


a
Ostreococcus tauri RCC4221
11265 aa
Ostreococcus tauri RCC4221 'Exeter'
19166 aa
Ostreococcus tauri RCC1115/RCC7079/7080/7081/7082
7282 aa
Ostreococcus tauri RCC7083
4794 aa
Ostreococcus lucimarinus CCE9901
EntF
18193 aa
Micromonas Commoda RCC299
14149 aa
Micromonas pusilla CCMP1545
7926 aa
Mantoniella tinhauana RCC11003
20889 aa
Keto reductase / Fatty acid synthase
Type 1 Polyketide Synthase
Adenylate forming domain
Enoyl Reductase
Phosphopantetheine attachment site
Sulfotransferase
b
Log2 abundance
15.0
12.5
Mycolactone [M+K-2H]-
Ebelactone A [M+H-2H2O]+
Ebelactone A [M-H2O-H]-
Actinoallolide A [M+CH3OH+H]+
6-Deoxyerythronolide B [M+H-2H2O]+
3 1 2 1 2 3
Ot
Ol
c
Mycolactone
Ebelactone A
6-Deoxyerythronolide B
Actinoallolide A