# 3D Genome Organization in Brown Algae

**Dissertation**

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Pengfei Liu

aus Henan, China

Tübingen

2026

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

| | |
|---|---|
| Tag der mündlichen Qualifikation: | 27.03.2026 |
| Dekan: | Prof. Dr. Thilo Stehle |
| 1. Berichterstatter/-in: | Prof. Dr. Susana M Coelho |
| 2. Berichterstatter/-in: | Prof. Dr. Rosa Lozano-Durán |

# Acknowledgements

Time has flown; it's already been five years since I started my PhD at the beginning of 2021. After this long, challenging and fascinating journey through my PhD, I would like to extend my sincerest gratitude to everyone who helped me complete this thesis. I would like to begin by expressing my most sincere gratitude to my advisor, Prof. Dr. Susana M. Coelho, for her unwavering support and insightful critiques throughout this project. Her commitment to academic excellence and detail has significantly shaped the dissertation. I am also grateful to Prof. Dr. Chang Liu at Hohenheim University for being my co-advisor and his exceptional guidance, valuable advice and availability throughout this project. I am also very grateful to the members of my TAC: Dr. Agnieszka Lipinska, Prof. Dr. Marja Timmermans and Prof. Dr. Rosa Lozano-Durán for their constructive feedback and essential suggestions that made my work better. I would also like to thank the research support team at MPI, Jeanette, Sibylle, Susan and many others for helping me with my PhD. Next, I would like thank everyone at Coelho's lab, specifically, I would like to thank Heike for her administrative work, as well as Agnes and Morgane for managing the laboratory and keeping it clean and organized; I would like to thank Min for taking such good care of me, encouraging me and giving me energy, and teaching me all the laboratory tips; I thank Rémy, Kenny, Andrea, Doro, Anaga, Masa and Jumana for their help with the algae cultures; I thank Fabian and Josue for their help with the bioinformatics, I thank Rita for her careful reading of this thesis and her constructive feedback. Finally, I would like to thank Michael, Siva, Elena, Carole, Liping, Cecile, Alex, Kohei, Claudia and Daniel for their help with scientific discussions and general chat. I would like to thank Sodai, Romy, Sabrina, Jeromine, Erica, Jia Xuan, Pauline, Elisa, Maria, Meri and many others for their friendship.
Last but not least, I would like to thank my family and friends for their emotional support. Specifically, I would like to thank my wife, Ying, for her support and companionship.

# Table of Contents

# Summary

The three-dimensional (3D) organization of the genome plays a crucial role in gene regulation and epigenetic states. In animal genomes, core 3D genome features like topologically associated domains (TADs) and chromatin loops have been extensively characterized, whereas in plants such higher order structures are less pronounced. Beyond animals and plants, however, relatively little is known about the 3D genome organization evolution.

This thesis explores the evolution of 3D genome organization in brown algae by comparing species differing in genome size, sexual systems (dioicous and monoicous), and morphological complexity. By substantially improving chromosomal-level assemblies, I demonstrated that brown algal genomes exhibit a high degree of linear conservation relative to the outgroup species, with only moderate rates of interchromosomal rearrangement. Examining interspecies variations in 3D chromatin organization revealed distinct patterns of nuclear architecture in brown algae shaped by evolutionary history. Comparative analysis suggests that 3D genome architecture is broadly conserved over approximately 11 million years of brown algae evolution. Notably, sex chromosomes and sex homologs display divergent interactions profiles distinct from autosomes. Consistent non-Rabl chromosome folding patterns were identified, which enabled the characterization of “centromere interaction clusters” and “telomere interaction clusters”. Identification of the centromeric regions revealed long-term co-evolution with a specific lineage of “centrophilic” LTR retrotransposons, which are uniquely present at the centromeres in most brown algal species and likely constitute the centromeric repeat.

This study presents the first comparative analysis of 3D chromatin organization across six phylogenetically diverse brown algal species and one outgroup, offering new insights into 3D genome architecture within the eukaryotic lineage through the integration of high-resolution Hi-C data, gene expression profiling, centromere characterization, and ancestral genome reconstruction.

# Zusammenfassung

Die dreidimensionale (3D) Organisation des Genoms spielt eine zentrale Rolle in der Genregulation und den epigenetischen Zuständen. Während in Tiergenomen zentrale 3D-Strukturen wie topologisch assoziierte Domänen (TADs) und Chromatin-Loops umfassend beschrieben sind, sind derartige höhergeordnete Strukturen bei Pflanzen weniger ausgeprägt. Jenseits von Tieren und Pflanzen ist jedoch nur wenig über die Evolution der 3D-Genomorganisation bekannt.

In dieser Arbeit wird die Evolution der 3D-Genomorganisation in Braunalgen untersucht, indem Arten verglichen werden, die sich in Genomgröße, Sexualsystem (diözisch versus monoizisch) und morphologischer Komplexität unterscheiden. Durch die Verbesserung der chromosomalen Referenzassemblierungen konnte gezeigt werden, dass Braunalgen-Genome eine hohe lineare Konservierung gegenüber Außengruppenarten aufweisen, mit lediglich moderaten Raten interchromosomaler Umlagerungen. Inter-artspezifische Unterschiede in der 3D-Chromatinorganisation offenbaren unterschiedliche Muster der Nukleararchitektur, die stark von der evolutionären Geschichte geprägt sind.

Vergleichende Analysen legen nahe, dass die 3D-Genomarchitektur über etwa 11 Millionen Jahre der Braunalgen-Evolution weitgehend konserviert geblieben ist. Geschlechtschromosomen und ihre homologen Regionen weisen dabei divergente Interaktionsprofile auf, die sich deutlich von den Autosomen unterscheiden. Konsistente nicht-Rabl-Chromosomenfaltungsmuster ermöglichten die Identifizierung von „Centromer-Interaktionsclustern“ und „Telomer-Interaktionsclustern“. Die Analyse der centromerischen Regionen zeigte eine langfristige Koevolution mit einer spezifischen Linie „centrophiler“ LTR-Retrotransposons, die in den Centromeren der meisten Braunalgenarten vorkommen und vermutlich die centromerischen Wiederholungen bilden.

Diese Studie liefert die erste vergleichende Analyse der 3D-Chromatinorganisation in sechs phylogenetisch unterschiedlichen Braunalgenarten und einer Außengruppe und bietet neue Einblicke in die 3D-Genomarchitektur eukaryotischer Genome durch die

Integration hochauflösender Hi-C-Daten, Genexpressionsprofile, Centromercharakterisierung und Rekonstruktion des Vorfahrengenoms.

# List of publications

† Shared first authorship

1. **Pengfei Liu**, Jeromine Vigneau, Rory J. Craig, Josue Barrera-Redondo, Elena Avdievich, Claudia Martinho, Michael Borg, Fabian B. Haas, Chang Liu, and Susana M Coelho. 2024. **3D chromatin maps of a brown alga reveal U/V sex chromosome spatial organization.** *Nature Communications*, 15, 9590. [DOI: https://doi.org/10.1038/s41467-024-53453-5]

2. Josué Barrera-Redondo†, Agnieszka P Lipinska†, **Pengfei Liu,** Erica Dinatale, Guillaume Cossard, Kenny Bogaert, Masakazu Hoshino, Komlan Avia, Goncalo Leiria, Elena Avdievich, Daniel Liesner, Rémy Luthringer, Olivier Godfroy, Svenja Heesch, Zofia Nehr, Loraine Brillet-Guéguen, Akira F. Peters, Galice Hoarau, Gareth Pearson, Jean-Marc Aury, Patrick Wincker, France Denoeud, J Mark Cock, Fabian B. Haas, and Susana M Coelho. 2025. **Origin and evolutionary trajectories of brown algal sex chromosomes.** *Nature Ecology and Evolution* [DOI: https://doi.org/10.1038/s41559-025-02838-w]

3. Jeromine Vigneau†, Jaruwatana Sodai Lotharukpong†, **Pengfei Liu**, Remy Luthringer, Bérangère Lombard, Damarys Loew, Fabian Haas, Michael Borg, Susana M Coelho. 2025. **Rewiring of chromatin regulation underlies the evolution of brown algal multicellularity**. *bioRxiv* (in revision) [DOI: https://doi.org/10.1101/2025.09.16.676480]

4. **Pengfei Liu,** Rory J. Craig, Elena Avdievich, Fabian B. Haas, Chang Liu, Susana M Coelho. 2025. **Lineage-wide evolution of 3D genome organisation and centromeres in brown algae**. *bioRxiv* (in revision) [DOI: https://doi.org/10.64898/2025.12.07.692804]

# 1 Introduction

## 1.1 The discovery of genome architecture in the nucleus

How genetic material is organized so compactly within the confined space of a cell or nucleus is a longstanding question in cell biology[1]. Following the discovery of chromosomes in the late 19th century, Walther Flemming first described mitosis in 1878 and noted that chromosomes exist as distinct entities[2]. Shortly thereafter, in 1885, Carl Rabl proposed that chromosomes retain a polarized arrangement in interphase nuclei, with centromeres and telomeres positioned toward opposite poles, a configuration later known as the Rabl configuration[3]. In 1909, Theodor Boveri further suggested that each chromosome occupies a distinct nuclear region, termed “chromosome territory”[4]. This hypothesis remained controversial for decades because interphase chromatin appeared diffuse under light microscopy. This long-standing controversy was eventually resolved by advances in the field of electron microscopy and biochemistry in the mid-20th century, which first revealed higher-order packaging of DNA, including the discovery of nucleosomes in 1974 (Roger Kornberg)[5]. Models such as the 30 nm fiber, which is formed by helical nucleosome folding and contains approximately 147 base pairs of DNA coiled around histone proteins, two copies of each H2A, H2B, H3 and H4, were subsequently proposed to explain additional levels of chromatin[6,7]. Later, in the 1980s~1990s, fluorescence in situ hybridization (FISH)[8] provided direct evidence that interphase chromosomes are organized into discrete territories, lending support to Boveri’s earlier hypothesis and shifting attention to the functional consequences of nuclear architecture for gene regulation. A major breakthrough in the field came with the development of chromosome conformation capture (3C) based approaches, culminating in the Hi-C technique[9], a technology based on proximity-ligation and high throughput sequencing that captures genome-wide physical interactions between chromatin regions (**Fig. 1**). Hi-C analyses identified large-scale compartmentalization (A/B compartments), as well as DNA looping interactions between regulatory elements. Subsequent work revealed the existence of topologically associating domains (TADs)[10] and other hierarchical features of genome folding. Together, these findings established chromosome folding as a central principle of nuclear organization and a key determinant of genome function.

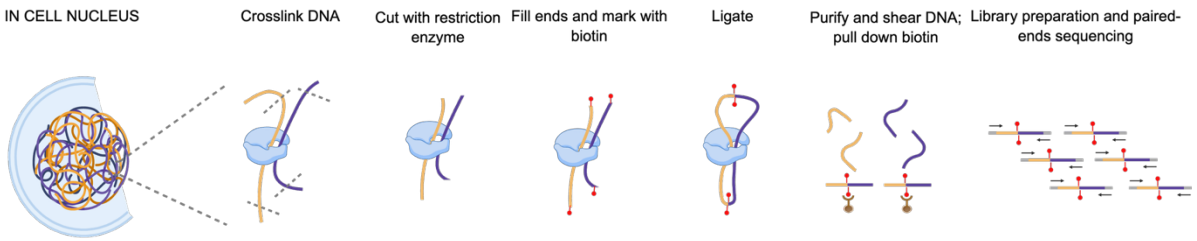



Fig. 1. Workflow of *in situ* Hi-C experiment.

A standard *in situ* Hi-C protocol involves the following steps: chromatin is first cross-linked to preserve spatial DNA interactions, followed by nuclei isolation. Then the digested DNA ends are filled in with biotinylated nucleotides. Proximal DNA fragments are subsequently ligated within intact nuclei, capturing chromatin contacts. Finally, cross-links are reversed, DNA is purified, and biotin-labeled ligation products are enriched and used for Illumina library preparation and high-throughput sequencing.

## 1.2 Hierarchical folding of chromatin

Eukaryotic genomes are hierarchically packed into multiscale structural units within the nucleus, ranging from large to small scales and including chromosomes territories (CTs), A/B compartments, TADs and chromatin loops[11–13](**Fig. 2**). This hierarchical genome organization enables precise transcriptional control of genes[14,15], DNA replication, and cellular differentiation[16]. Advances in high-throughput chromosome conformation capture (e.g., 3C, 4C, 5C, Hi-C, and Micro-C),[9,15,17,18] together with high-resolution microscopy imaging technology, have allowed the characterization of these multiscale structures directly and the comparison of 3D genome organization across different species[9,19–21].

At the global level, interphase chromosomes adopt distinct spatial configurations and comprise two different types of chromosomal architectures. In type I (Rabl / Rabl-like) organization, centromeres and/or telomeres cluster together, whereas type II organization features discrete chromosome territories with strong interactions between loci on the same chromosome[22]. Within these broad genome folding structures, the genome is further partitioned into A/B compartments. The A compartments are gene-rich and contain high-GC content, corresponding to transcriptionally active chromatin domains located in the inner part of nucleus; in contrast, B compartments are repeat-rich and contain low-

GC content, corresponding to transcriptionally inactive chromatin domains located at the periphery of nucleus[9,23].

At finer scales, the genome is organized into TADs, TADs are self-interacting regions that are insulated from neighboring chromatin and typically range from tens of kilobases to several megabases in genomes[10,24]. Additionally, and at the most local level, chromatin loops can be recognized by prominent contacts between two genomic loci, and play a critical role in regulating regulatory elements interactions, facilitating precise gene regulation and other functional elements[1,13].

Together, these hierarchical layers, from chromosome territories to chromatin loops, establish the 3D genome framework that coordinates gene regulation and genome function in eukaryotic cells.

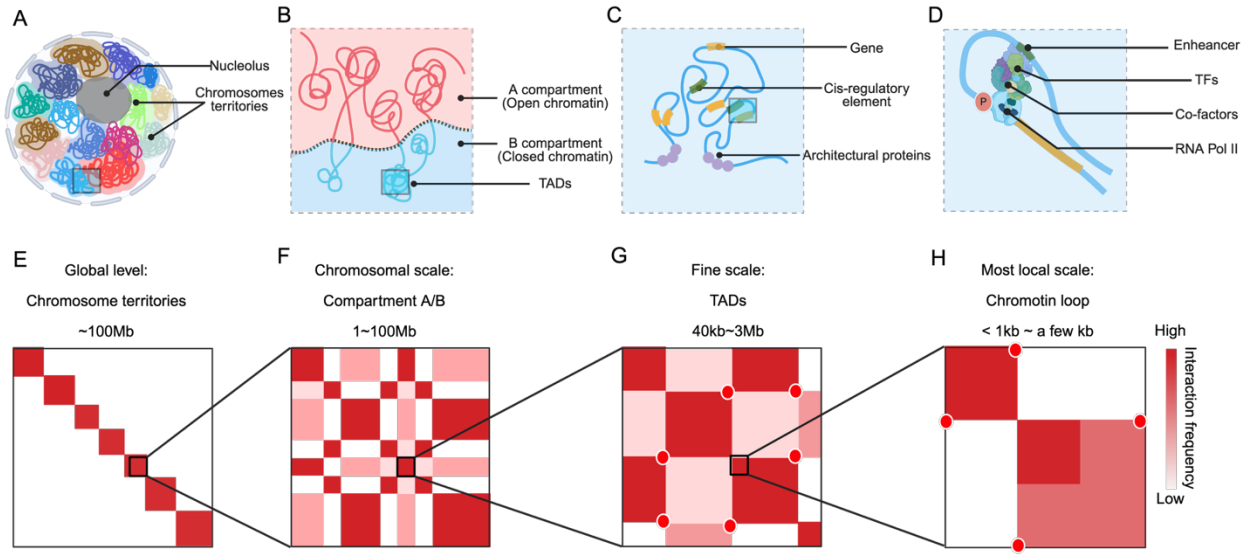



Fig. 2. Hierarchical 3D genome organization detected by Hi-C.

(A) During interphase, individual chromosomes occupy discrete nuclear regions, referred to as chromosome territories. At global level, in Hi-C contact matrices, these territories appear as square blocks along the diagonal, reflecting enriched intrachromosomal (cis) interactions relative to interchromosomal (trans) interactions (E).

(B) At the chromosomal scale, genomes can be partitioned into large-scale chromatin compartments with preferential interactions with itself, often corresponding to relatively open, transcriptionally active regions and more compact, transcriptionally inactive regions. These compartments are visualized as plaid patterns in normalized Hi-C contact maps (F).

(C, G) At fine scales, chromatin can be organized into self-interacting domains, commonly referred to as topologically associating domains (TADs, in animal) or TAD-like domains (in plant), which exhibit elevated interaction frequencies within domains and reduced interactions across domain boundaries. The molecular features associated with domain boundaries vary across taxa and cell types.
(D) At most local scale, specific chromatin loops connect distant genomic loci into close in 3D space, facilitating regulatory interactions. In Hi-C matrices, such loops are detected as localized, high-intensity interaction signals (H), although their prevalence and underlying mechanisms differ among organisms.

## 1.3 Conserved and divergent 3D genome organization across eukaryotes

Since the first Hi-C experiment characterized the 3D genome organization in human cell lines[9], this technology has been widely applied to diverse eukaryotic genomes to investigate the principles of genome architecture across the tree of life[11,13]. Comparative Hi-C studies have uncovered remarkable variability in genome organization across taxa[9,18,21,25–28], spanning scales from individual genes to entire chromosomes, and have revealed both conserved and lineage-specific features of higher-order chromatin folding. These findings raise a fundamental question: is there a common folding principle governing 3D genome architecture across eukaryotes?

### 1.3.1 Chromosome territories

A major feature of nuclear organization is the formation of CTs, in which each chromosome occupies a discrete and largely non-overlapping spatial domain within the interphase nucleus[9]. CTs have been described across all eukaryotic lineages and are considered functionally relevant for genome evolution[29]. The spatial arrangement of CTs is not random, it influences long-range DNA–DNA interactions, heterochromatic gene silencing, genome stability, DNA repair, and other nuclear processes[30]. Despite extensive characterization, the molecular determinants that specify CTs positioning between different chromosomes remain poorly understood[29,31].
A recent comparative study of 3D genome folding across 24 species (including animals, fungi, and plants), identified two types of nuclear architecture based on Hi-C data[22]. The first corresponds to canonical chromosome territories, whereas the second is

characterized by frequent centromeres and/or telomeres contacts that polarize chromosomes toward opposite nuclear poles, known as the Rabl / Rabl-like configuration. Remarkably, a single factor, the presence or absence of condensin II subunits, was suggested to distinguish these two architecture types[22]: species displaying a predominant Rabl/Rabl-like configuration had lost condensin II, whereas species that formed chromosome territories retained it. In contrast, another study reported that species with a fully functional condensin II complex, such as Drosophila, exhibit a pronounced Rabl/Rabl-like configuration, with centromeres and telomeres segregated to opposite poles of the nucleus[32]. These apparent discrepancies raise the question of what role condensin II actually plays in centromere and/or telomere organization in systems that follow the Rabl/Rabl-like paradigm.

Plant genomes, which vary widely in both genome size and transposable element (TE) content, display an especially broad range of nuclear architectures [12]. For example, the relatively small genome of *Arabidopsis thaliana* contains repeats enriched at pericentromeric regions that cluster into densely packed chromocenters, while euchromatic chromosome arms extend outward, giving rise to the characteristic “Rosette” configuration [33]; In contrast, species with much larger genomes such as rice[34], wheat[35] tomato, sorghum[36] and lily[37], often display a telomere bouquet, in which telomeres cluster within a restricted region of the nucleus. These distinct nuclear topologies provide substantial flexibility in both local and global chromatin interactions and likely facilitate adaptive regulatory responses to developmental and environmental cues[32,38], as well as homologue chromosome recombination and segregation[39].

### 1.3.2 A/B compartments

Beyond the level of CTs, interphase chromosomes are further divided into A and B compartments, referring to transcriptionally active (A) and inactive (B) chromatin, respectively. A/B compartments are reflected as plaid or checkerboard patterns in Hi-C contact maps. The A compartments show strong interaction frequency with other A compartments along and between chromosomes, and similar preferential interactions occur among B compartments. This phenomenon is referred to as

compartmentalization[1,9]. Compartmentalization represents a secondary, yet evolutionarily conserved, layer of genome organization. In mammalians, early low-resolution Hi-C data revealed that the genome can be partitioned into two broad types of compartments corresponding to euchromatin and heterochromatin[9]; subsequent high-resolution Hi-C analyses demonstrated that these compartments can be further subdivided into sub-compartments, with kilobase-scale A compartments showing precise co-localization with active histone marks and highly expressed genes[40]. In plants, A/B compartments have been identified for many species with different genome sizes[12,41]. In general, the A compartment is formed by euchromatin chromosome arms, while the B compartment is formed by pericentromeric heterochromatin and centromeres[36]. In small plant genomes, like the *A. thaliana* genome, compact and loose structural domains are defined by interaction frequency, and correlate with epigenetic chromatin states[18,42]. In large plant genomes, additional local A and B compartments can be found in highly gene-rich islands inside heterochromatin and TE-rich euchromatin on individual chromosomes[12,36,41], and strongly correlate with DNA replication timing[35,43].

Compartments have been identified in the genomes of most eukaryotic organisms, suggesting an essential role in chromatin maintenance and genome stability. Notably, compartment identity is dynamic and can switch during cell differentiation[44], development[45] or evolution[20,46,47], for example, approximately 36% of compartments undergo state transitions in human cells[44], while switched A/B compartments were also identified in the development from vegetative nuclei to sperm nuclei during male gametogenesis in *A. thaliana*[45]. So far, phase separation-mediated interactions between heterochromatin have been proposed to establish compartmentalization[48], but the molecular mechanisms for the formation of compartments is not known[23].

### 1.3.3 TADs and chromatin loops

Each compartment comprises multiple TADs, which are structural units typically spanning hundreds of kilobases to several megabases[49]. In vertebrates, structural maintenance of chromosomes complexes (SMC) include cohesin and condensin, and performs an essential function in shaping the structure of TADs[50]. In Hi-C data, TADs exhibit strong

intra-domain contact frequency[49] and their boundaries are often enriched in highly expressed features (eg., transfer RNA, housekeeping genes, and active chromatin marks)[51]. These boundaries are often stabilized by architectural proteins and specific DNA motifs to mediate long-range interactions[52]. In the human genome, the insulator protein CTCF[10] can define TAD boundaries and maintain chromatin loop structures[53], whereas in *Drosophila* various insulator proteins colocalize at TADs boundaries, with most boundaries being bound by one or more insulator proteins[54]. This suggests co-binding of insulator proteins is essential for TADs boundary formation in *Drosophila*.

Unlike animals, most plant genomes lack canonical TADs mainly due to the absence of CTCF[50], and rather display a more diverse TAD-like organization. Several studies on plant chromatin regulators have advanced our understanding on the mechanism of regulation and maintenance of TAD-like structure in plant genome[41]. For example, in *Arabidopsis* EMBRYONIC FLOWER1 (EMF1) was shown to act as genome modulator by interacting with SISTER CHROMATIN COHESION3 (SCC3) at domain boundaries to regulate chromatin loop formation[55], while TAD-like regions can be formed by BMI1 and condensin complexes[56]. Additionally PDS5A contributes to TAD formation without gene expression and histone modifications changes [57]. In rice, the Microrchidia (MORC) 6b protein found to interact with PRC2 at domain boundaries[58]. Collectively, in plants the presence of TAD-like structures correlates with regulatory features and histone modification patterns[13,24], and is also largely shaped by gene transcription [18].

TADs are evolutionarily conserved within close related lineages and cell types, and are considered fundamental units of chromosome folding[20,59,60]. For example, 30-40% of TADs are conserved over ~49 million years (Myr) of *Drosophila* genome evolution[27]. Despite this robustness, 3D genome organization must retain sufficient flexibility to permit dynamic chromatin remodeling during processes such as transcriptional reprogramming, DNA replication, and mitosis[53,61].

Chromatin loops, a fundamental structural unit of the eukaryotic genome, bring distal regulatory elements into proximity. These loops are primarily formed through the cohesin-mediated loop extrusion model[62]. In this process, the cohesin complex is first loaded onto the chromatin fiber at specific sites. Once loaded, it begins to extrude the DNA through its ring structure. Extrusion continues until cohesin encounters a pair of CTCF proteins

bound to DNA in a convergent orientation, which stalls the complex[10]. This stabilized complex, pinned between two CTCF sites, defines the base of a chromatin loop and brings two distant genomic loci into close proximity. These loops represent prominent interactions in Hi-C maps and are relatively conserved across lineages during evolution. For example, Anopheles genomes feature specific, extremely long-range chromatin loops, spanning up to 31 Mbp, and conserved for ~100 Myr[20]. Since CTCF plays a critical role in loop extrusion, in the absence of this protein, loop interactions are generally weaker and require high-resolution or regional capture-based Hi-C data for detection[53]. Recent research indicate that the chromatin loop is the ancestor genome architecture in the animal lineage by comparing with the last unicellular relative organism[63].

Collectively, these findings highlight both the conserved and lineage-specific features of 3D genome organization across eukaryotes, providing a conceptual framework for exploring how higher-order chromatin folding, including A/B compartmentalization and TADs are established, maintained, and diversified during genome evolution of eukaryotic lineages.

To date, chromatin architecture has been extensively characterized in animals[63–66] and plants[18,26,36,67], revealing that 3D genome organization is shaped by factors such as genome duplications[68], genome size[69,70], TEs[71,72], and gene content[73]. Comparative Hi-C analysis, which integrate chromatin contact maps across species, have provided powerful insights into how 3D genome organization has been shaped by evolution[13] and how those structures contribute to developmental regulation[20,45,46,59,74,75] and speciation research[76]. Together these data allow rigorous characterization of the conservation and divergence of 3D genome architectures across species, even those separated by significant evolutionary distances[20,28,59,63].

## 1.4 The brown algae

Brown algae are an evolutionary lineage distinct from animals and plants, with common ancestors dating back more than more than a billion years[77], and have an immense ecological importance[78,79]. Brown algae have more than 2000 described species, display a diversity of genome sizes (from ~200 Mbp to ~1.2 Gbp), sexual systems (U/V, XX/XY

and hermaphrodite), and morphological complexity[80]. They are interesting *per se* as representatives of the eukaryotic biodiversity, however, they are also increasingly being used as a model for comparison in genomic, developmental and evolutionary studies [81–84]. Specifically, they represent an interesting model to decipher the history of sex chromosomes due to their diversity of reproductive systems, types of life cycle and sexuality systems[78,81]. Phylogenetic analyses indicate a frequent cycling through different feature variants across evolutionary history, offering a unique opportunity to explore how sex chromosomes evolve in organisms with differing life-history traits[84]. An additional advantage of the brown algae as a model group is the availability of genomic information for a considerable number of species, such as several *Ectocarpus*, Desmarestiales, and kelps[85]. These species now boast a number of genetic and genomic resources, including high quality genomes, transcriptomes and genetic and physical maps [80,85,86]. Interestingly, this collection of brown algae include species that have evolved towards hermaphroditism from dioicous ancestors (separate sexes, with sex chromosomes), as well as those that have evolved in the opposite direction, with dioicous species arising from hermaphrodite ancestors at the basis of the brown alga phylogeny[87].

Three principle types of sex chromosome systems have been described across eukaryotes: the XX/XY sex determination system found in mammals[88] and most plants, the ZW/ZZ sex determination system characteristic of birds and butterflies[89]. However, many eukaryotes have UV sex determination systems, such as those of mosses, fungi and some algae[90] and the evolutionary dynamic and structural organization of sex chromosomes and sex determination in these systems is not fully understood[91,92].

Presently, we still lack a comprehensive understanding of the chromatin landscapes and 3D chromatin organization of sex chromosomes, and how their folding dynamics underpin sex-specific gene expression and sexual differentiation, and the transition from dioicy to monoicy in brown algae. Comparative analysis of 3D genome organization across these different species will provide valuable insights not only into the types and mechanisms of genome organization in this understudied group of eukaryotes, but also into how sex chromosome structure varies with sexual dimorphism, and the evolutionary history of sex chromosomes in a recombining context in hermaphrodites.

# 2 Objectives

My PhD project aims to uncover how 3D genome organization has evolved in brown algae. To accomplish this, I selected six representative brown algal species with varying genome sizes, sexual systems (including dioicous species with separate male and female strains, and monoicous or hermaphroditic species) and a relevant outgroup species. I optimized plant *in situ* Hi-C protocol and generated several new reference genomes for brown algae species. Using these data, I investigated structural features of 3D genome architectures in brown algae. I further examined whether and how the 3D architecture of sex chromosomes has changed during the evolutionary transition from a 'simple' sex chromosome with a small non-recombining SDR to a highly heterochromatic and largely non-recombining sex chromosome. In parallel, I analyzed hermaphroditic species to explore chromatin remodeling with the transition from separate sexes to hermaphrodites. Together, these analyses shed light on how sex chromosomes are spatially organized within the 3D nucleus and how this organization correlates with the diversity of reproductive and sex-determination systems observed in brown algae.

# 3 Summary of results

## 3.1 3D chromatin maps of a brown alga reveal U/V sex chromosome spatial organization

**Pengfei Liu**, Jeromine Vigneau, Rory J. Craig, Josue Barrera-Redondo, Elena Avdievich, Claudia Martinho, Michael Borg, Fabian B. Haas, Chang Liu, and Susana M Coelho. 2024. *Nature Communications*, 15, 9590.
[DOI: https://doi.org/10.1038/s41467-024-53453-5]

### 3.1.1 Synopsis

The understanding of 3D genome organization within the nucleus remains surprisingly limited, even in classical model systems. We generated a near complete reference genome of the model brown alga *Ectocarpus* combined with high resolution Hi-C maps. We found that *Ectocarpus* interphase chromatin adopts a non-Rabl configuration, with telomeres and centromeres of all 27 chromosomes forming clusters within the 3D nuclear space. The chromatin exhibits a highly streamlined 3D architecture, lacking canonical TADs, and A/B compartments are largely defined by H3K79me2 deposition. Focusing on the U and V sex chromosomes, we observed that the sex-determining region spans the centromere in both cases and is strongly insulated from the remainder of the chromosome. Whilst there were overall no differences in the 3D chromatin organization between male (V) and female (U) chromosomes, the nuclear 3D chromatin organization of sex chromosomes, specifically within the non-recombining region, is markedly distinct compared with autosomes. Remarkably, *Ectocarpus* centromeres feature a centromere-specific retrotransposon that encodes a chromodomain, implying evolutionary convergence of centromere-specific integration of TEs in brown algae and land plants, despite a billion years of independent evolution. Finally, we revealed the unique 3D chromatin architecture of a chromosome 6 region containing a giant endogenous viral

element, highlighting the interplay between giant dsDNA viruses and host chromatin and providing initial insights into their potential functional role.

### 3.1.2 Own contribution

First author and main investigator. I cultured the *Ectocarpus* male and female strains; I optimized protocols and constructed high quality Hi-C libraries and generated the final genome assembly; I performed the 3D genome analysis under the supervision of Chang Liu and Susana M Coelho; Elena Avdievich generated Nanopore long reads; Fabian B. Haas did the initial genome assembly; Rory J. Craig characterised *Ectocarpus* centromeres; Rory J. Craig, Jeromine Vigneau and Michael Borg profiled chromatin states of centromeres; Eventually I wrote the manuscript together with Rory J. Craig, Chang Liu, Susana M Coelho with input from Josue Barrera-Redondo, Claudia Martinho, Michael Borg and other co-authors.

## 3.2 Origin and evolutionary trajectories of brown algal sex chromosomes

Josué Barrera-Redondo†, Agnieszka P Lipinska†, **Pengfei Liu,** Erica Dinatale, Guillaume Cossard, Kenny Bogaert, Masakazu Hoshino, Komlan Avia, Goncalo Leiria, Elena Avdievich, Daniel Liesner, Rémy Luthringer, Olivier Godfroy, Svenja Heesch, Zofia Nehr, Loraine Brillet-Guéguen, Akira F. Peters, Galice Hoarau, Gareth Pearson, Jean-Marc Aury, Patrick Wincker, France Denoeud, J Mark Cock, Fabian B. Haas, and Susana M Coelho. 2025. *Nature Ecology and Evolution*.
[DOI: https://doi.org/10.1038/s41559-025-02838-w]
† Indicates equal contribution

### 3.2.1 Synopsis

Knowledge of sex chromosome biology and evolution largely stem from studies of diploid systems, whereas the study of sex chromosomes in the context of haploid systems remains comparatively underdeveloped. Brown algae represent a distinctive lineage to investigate sex chromosome evolution, as they possess a U/V system that is fundamentally different from the diploid sex determination (XY and ZW) systems of animals and plants, providing a valuable model for elucidating the evolutionary history and diversification of sex chromosomes. In this study, we examined nine representative brown algal species encompassing broad phylogenetic, morphological, and reproductive diversity, complemented by an outgroup species. Using the available genomic resources, we improved several assemblies to a level close to chromosome-scale. Utilizing these new genomes, we observed largely conserved karyotypes (27–33 chromosomes) and evolutionary conservation of sex chromosomes among dioicous lineages through macrosynteny analysis. Our data shows that the U/V sex chromosomes in brown algae appear to have originated between 450 and 224 million years ago (Mya), following the suppression of recombination around the male-determining gene (*MIN*) on the proto-U / -V chromosomes. Subsequent lineage-specific expansions of the sex-determining region (SDR) accompanied increasing morphological complexity and sexual dimorphism.

Our results also shows that brown algal U/V chromosomes follow an evolutionary trajectory characterized by gene gain rather than the degenerative pattern seen in XY and ZW systems, driven by nested inversions, TE accumulation, and *de novo* gene emergence. Gene traffic into and out of the SDR has further shaped its composition over more than 100 million years. Transitions from ancestral U/V to derived XY systems involved the relocation of V-linked genes and erosion of U/V-specific features, while hermaphroditism may have arisen through ectopic acquisition of U-linked genes in ancestral males.

### 3.2.2 Own contribution

Co-author. I validated SDR scaffolds using PCR; I cultured the *Desmarestia herbacea* male and female strains and constructed all the Hi-C libraries and preformed further data analysis.

## 3.3 Rewiring of chromatin regulation underlies the evolution of brown algal multicellularity

Jeromine Vigneau†, Jaruwatana Sodai Lotharukpong†, **Pengfei Liu**, Remy Luthringer, Bérangère Lombard, Damarys Loew, Fabian Haas, Michael Borg, Susana M Coelho. 2025. *bioRxiv.*
[DOI: https://doi.org/10.1101/2025.09.16.676480]
† Indicates equal contribution

### 3.3.1 Synopsis

Our current understanding of histone post-translational modifications and chromatin architecture is largely derived from a limited number of animal and plant lineages. In this study, we investigated the evolution of chromatin regulation within the brown algal lineage by integrating ChIP-seq, western blotting, and mass spectrometry analyses. Through comparative genomic and epigenomic approaches, we revealed that multicellular brown algae represent highly divergent eukaryotes in terms of chromatin regulation. Remarkably, the lack of canonical DNA methylation and heterochromatin-associated modifications like H3K27me3 appears to be an ancestral feature of this group, while H3K79me2, has adopted a repressive function. We show that chromatin states are conserved across the lineage: broadly conserved genes exhibit stable histone modification profiles, whereas lineage-specific or recently evolved genes are frequently embedded within repressive chromatin regions. Moreover, we characterize chromatin organization across the sex chromosomes, uncovering distinct epigenetic patterns between sex chromosomes. Finally, analysis of a sister lineage to brown algae revealed the retention of cytosine methylation, suggesting an evolutionary transition in epigenetic regulation accompanying multicellularity in this group.

### 3.3.2 Own contribution

Co-author. I helped developing protocols for ChIP-seq. I performed data exploration, analysis and visualization.

## 3.4 Lineage-wide evolution of 3D genome organisation and centromeres in brown algae

**Pengfei Liu,** Rory J. Craig, Elena Avdievich, Fabian B. Haas, Chang Liu, and Susana M Coelho. 2025. *bioRxiv.*
[DOI: https://doi.org/10.64898/2025.12.07.692804]

### 3.4.1 Synopsis

How 3D genome organization evolves across deep evolutionary timescales and diverse eukaryotic lineages remains poorly understood. Previously, I charactered 3D genome organization of model brown alga *Ectocarpus* sp.7, in this study, we present chromosome-level assemblies and a comprehensive comparison of genome structures across six representative brown algal species and one sister outgroup. Among these species, genome sizes range from approximately 200 Mbp to 851 Mbp, varied sexual systems are present (dioicous and monoicous), and different morphological complexity levels are observed, reflecting the lineage's phylogenetic diversity. We show that brown algal genomes exhibit a high degree of linear conservation relative to the outgroup, experiencing only a moderate rate of inter-chromosomal rearrangement events. We revealed that 3D genome organization of brown algae display distinct patterns of nuclear architecture shaped by evolutionary history: core features of chromatin compartmentalization, lineage-specific domain-level genome architecture, and 3D genome architecture is conserved at the scale of ~11 Myr of brown algae evolution. Notably, sex chromosomes and sex homologs display divergent chromatin interaction profiles compared to autosomes. Finally, we identified consistent non-Rabl chromosome folding patterns that enable characterization of "centromere clusters" and "telomere clusters". In addition, the identification of the centromeric regions revealed long-term co-evolution with a specific lineage of "centrophilic" LTR retrotransposons. These are uniquely present at the centromeres in most species and likely constitute the centromeric repeat. The centromeres are also hotspots for chromosomal rearrangements and we

characterized several Robertsonian translocations and centric insertions that have occurred in brown algal evolution.

### 3.4.2 Own contribution

First author and main investigator. I prepared all the algae cultures; I constructed high quality Hi-C libraries for each species and generated final genome assemblies; I performed the comparative 3D genome analysis under supervision of Chang Liu and Susana M Coelho; Elena Avdievich generated Nanopore long reads; Fabian B. Haas did the initial genome assemblies; Rory J. Craig characterized centromeres; I wrote the manuscript together with Susana M Coelho, Chang Liu, Rory J. Craig with input of Fabian B. Haas, Elena Avdievic.

# 4 Discussion

## 4.1 Scope of Genomic Resources

In this thesis, I optimized an *in situ* Hi-C protocol[26] for plant and adopt it to work with brown algae and I leveraged the power of Hi-C to generate proximity ligation data for several species in this group. Combining Oxford Nanopore sequencing and the Hi-C technique, I generated chromosome-scale, high-quality reference genomes for a total of eleven brown algal strains, belonging to seven different species. This comprehensive effort includes the *de novo* assembly of two genomes (the brown algae *U. pinnatifida* female, and *D. dichotoma* female) **(paper 3.4)** and the chromosomal-level upgrading of nine previously published assemblies (*Ectocarpus* sp.7 male and female **(paper 3.1)**, *C. linearis*, *U. pinnatifida* male, *D. herbacea* male and female, *D. dichotoma* male, and the sister outgroup species *S. ischiensis* **(paper 3.4)**. Collectively, these eleven high-quality brown algal reference genomes establish an important comparative resource for investigating the evolution of genome structure across the Phaeophyceae, while simultaneously laying the foundational groundwork for the detailed 3D chromatin organization analysis further presented in this work.

## 4.2 Macrosynteny and karyotype stability

Comparative analyses revealed uncovered extensive of macrosynteny conservation of the assembled brown algal species when compared with the outgroup species **(paper 3.2 and 3.4)**, despite their deep divergence time of approximately 450 Myr[93]. This unexpected stability suggests that the overall karyotype remains relatively stable across the Phaeophyceae **(paper 3.2 and 3.4)** [94,95], with chromosome numbers narrowly ranging from 27 to 32. Interchromosomal rearrangements were identified, especially in the Ectocarpales lineage **(paper 3.4)**, and their frequent occurrence in centromeric regions suggests that centromeres may play a role in driving evolutionary change as was hypothesized in *Drosophila* [96] and karyotype diversification through localized rearrangements as observed in *Malassezia* [97]. Smaller chromosomes may be more susceptible to chromosomal rearrangements (as seen in butterflies[64] and birds[98]), which can reconfigure higher-order folding by altering contact landscapes[60], shifting domain

boundaries, or modifying long-range interaction frequencies. Despite evidence of derived karyotypes in Ectocarpales, our analyses revealed a streamlined 3D genome architecture (CTs and compartment A/B), suggesting that this lineage follows a distinct mechanistic trajectory of chromosomal evolution within brown algae **(paper 3.4)**. Several hypotheses may explain this pattern: (i) selective pressures may favor a simplified folding regime that minimizes the functional consequences of rearrangements[11]; (ii) rapid life cycles and predominantly haploid stages may reduce the fitness costs of rearrangements[99], allowing karyotypic change without corresponding increases in architectural complexity; and (iii) genome compaction and repeat landscape features may bias chromosomes toward stable, low-complexity folding states[12]. Together, these mechanisms provide potential explanations for why extensive karyotypic evolution in Ectocarpales has not been accompanied by the emergence of more elaborate 3D genome organization. Further experiments will be needed to substantiate these hypotheses.

## 4.2 The 3D genome architecture of brown algal genomes

In this thesis, I present the first comparative analysis of brown algae 3D nuclear architecture in a multicellular eukaryote outside of the well-studied animal and land plant lineages. By investigating brown algal genomes, we uncovered fundamental features of 3D genome organization that are deeply conserved across Eukaryota. Key conserved structures of the 3D genome architectures in brown algae include the chromosome territories, non-Rabl configuration (the interaction clustering of centromere-to-centromere and telomere-to-telomere), and A/B compartments (segregation of active and inactive chromatin) **(paper 3.4)** [100]. The persistence of these structures, despite the phylogenetic distance from other eukaryotic supergroups, underscores their essential, non-negotiable roles as core organizational principles conserved across deep evolutionary timescales **(paper 3.4)**.

However, we also observed structural divergence linked to specific genomic parameters. Chromatin domains (or TAD-like) were detected in brown algal species with larger genomes **(paper 3.4)**, but were notably absent in smaller genomes like that of *Ectocarpus*[100]. Unlike animals, brown algae lack the key architectural protein CTCF (CCCTC-binding factor), which is a major driver of TADs stability and loop extrusion in

mammals[10,13]. In the small genome of *Ectocarpus*, the observed compartmentalization appears to be driven primarily by differential histone modifications **(paper 3.1)**[100]. This absence of chromatin domains suggests a flexible, perhaps more environmentally responsive, mechanism for gene regulation, relying on dynamic chromatin marking rather than on static structural boundaries[1]. Understanding the correlation between specific genomic changes (like genome size, TE content) and chromatin architecture is essential for unraveling the molecular mechanisms through which genome evolution contributes to phenotypic diversity and speciation in this ecologically crucial group[11,101]. The primary difference between the disparate genome sizes observed in the analyzed genomes is the content of TEs **(paper 3.2 and 3.4)**. Consistent with this, the outgroup species, *S. ischiensis*, exhibits unique inter-chromosomal interactions specifically linked to particular repeat elements. These findings collectively indicate that TEs likely play a crucial, lineage-specific role in shaping 3D genome organization in brown algae **(paper 3.4)**.

However, although a high-quality, curated TE library is available for *Ectocarpus*[102], TE annotation for the other species analyzed here remains technically challenging. Future work involving comprehensive manual curation and the development of lineage-specific TE libraries will be essential to determine which TE families may contribute to or reinforce higher-order chromatin architecture.

## 4.3 Brown algae centromere evolution

Centromeres play a crucial role in assembling kinetochore protein complexes and anchoring spindle microtubules, thereby ensuring the accurate segregation of chromosomes during cell division[103,104]. Despite this critical and conserved cellular function in sister chromatid segregation, the underlying DNA sequences and associated proteins evolve rapidly across species, which presents a paradox in genetics and evolutionary biology[105]. This rapid sequence evolution is consistent with the epigenetic nature of centromere identity and with the tendency of centromeres to accumulate satellites, tandem repeat arrays, and interspersed TEs. The deposition of the histone H3 variant CENH3 (also known as CENP-A) defines the functional centromeric domain, rather than any specific underlying DNA sequence. In many plants, centromeric satellites are densely packaged and heavily methylated, both at cytosines methylation and through

heterochromatic histone modifications, rendering them relatively inaccessible[106]. Such repressive chromatin environments are frequently invaded by centromeric retrotransposons, and in several species, these retrotransposons can even serve as preferential CENH3 loading sites, effectively becoming an integral component of the functional centromere[103]. We showed the model brown algae species *Ectocarpus* has a monocentromere, which displays a centromere-to-centromere cluster in Hi-C maps, as do all the other analysed species of the Phaeophyceae group **(paper 3.1 and 3.4)**. *Ectocarpus* monocentromeres feature a centrophilic retrotransposon: *Metaviridae* LTR (long terminal repeat retrotransposon), we termed as *ECR* (*Ectocarpus Centromere Retrotransposon*) element **(paper 3.1)**. Similar to centrophilic retrotransposons found in plants[103,107], we found a chromodomain at the C-termini of *ECR*. The presence of this chromodomain is significant as it likely enables retrotransposons to recognize and integrate specifically into heterochromatin by binding to specific histone modifications[107]. We hypothesize this mechanism also drives the centromere-targeted integration observed in *Ectocarpus*. However, we do not yet know precisely how the *ECR* chromodomain targets the centromere. The centromere-specific histone H3 variant CENH3 could be recognized, or else there could be binding to other centromere-associated proteins. Future research should include testing the binding specificity of the *ECR* chromodomain.

Homologs of *ECR* were initially found within the broad centromeric clusters defined from the Hi-C maps in each brown algal genome *ECRs* **(paper 3.4)**. However, we identified several other members of this lineage (*ECR* and/or *ECR-like*) across the brown algal genomes, including both centrophilic and non-centrophilic families by conducting a phylogenetic analysis of the combined Gag and Pol protein sequences of all identified *ECRs* **(paper 3.4)**. Our phylogenetic analysis of these groups of centrophilic and non-centrophilic elements show that centrophily evolved from a non-centrophilic LTR retrotransposon, and centrophily has appeared in the common ancestor of the brown algal crown radiation (BACR) at least 160 Myr ago[93]. Centromere enrichment of specific LTR retrotransposon lineages is not unique to brown algae. In grasses, for example, Ty3-gypsy-related centromeric retrotransposons (CRM/CRR/cereba families) are major and often CENH3-associated components of centromeres in maize, rice, barley and related

species[108,109]. These elements form distinct subfamilies with differing chromosomal distributions and appear to have been repeatedly recruited to centromeric chromatin during grass evolution[110]. More broadly, centromeres in many eukaryotes show heterogeneous mixtures of satellite repeats and transposable elements[111]; in *Drosophila*, for instance, centromeric regions are composed of unique combinations of TEs families rather than a single conserved element type, pointing to multiple independent events of TEs recruitment to centromeres across lineages[112].

Taken together, our results support a model in which diverse LTR retrotransposon families can be independently co-opted to centromeric function in different eukaryotic clades, and that the centrophilic *ECR*s in brown algae represent one such ancient recruitment event. This convergence suggests that sequence-specific targeting or chromatin-binding properties of certain retrotransposon integrases (or associated proteins) may predispose particular lineages to centromeric localization, and that subsequent diversification of those elements contributes to centromere structure and evolution[113]. Future work combining improved centromere-targeted ChIP strategies with curated TEs annotations will be essential for directly linking centrophilic retrotransposons to functional centromere identity and evolution in brown algae.

## 4.4 Divergence of sex chromosomes in 3D space

Sex chromosomes have a distinct chromatin organization, including unique local heterochromatic epigenetic landscapes, when compare to autosomes[88]. Highly differentiated sex determination regions (SDRs) on chromosomes have often evolved constitutive heterochromatin owing to TEs accumulation and have become associated with the chromocenter[114].

The majority of brown algae species exhibit a haploid phase U/V sex-determination system[91], and the U/V sex chromosomes in brown algae are inferred to have arisen between 450 and 224 Mya (**paper 3.2**). In this thesis, we found that despite sex chromosomes of brown algae being extremely old (ancestral), sex chromosomes across the brown algae group are not syntenic. This lack of synteny is largely related to changes in TE content, resulting in very large SDRs (e.g. *D. herbacea* or *U. pinnatifida*) and very

small ones (e.g. *Ectocarpus* sp. 7) **(paper 3.2)**. SDRs of sex chromosomes are spanning the centromeres, and are heterochromatic region compare with PARs (pseudoautosomal regions) or autosomes **(paper 3.4)**. Hi-C analysis revealed critical and previously unknown genomic interactions of sex chromosome: SDRs are predominantly annotated as A compartments, highly spatially insulated from PARs and the rest of the genome **(paper 3.4)**. This compartmentalization potentially facilitates the independent evolution of sex-specific traits by creating an optimized local chromatin environment (e.g., specific epigenetic marks or localized gene regulation)[88] that allows the sex chromosome linked genes to evolve separately from their autosomal counterparts. The highly repetitive nature of the centromeric regions, which often associates with the nuclear periphery and heterochromatin, likely plays an active role in establishing and maintaining this segregated architecture. Thus, this physical insulation likely serves as a powerful mechanism accelerating the evolutionary trajectory from ancestral autosomes to distinct, highly differentiated sex chromosomes. This finding provides unprecedented insight into the physical mechanisms maintaining sex chromosome integrity and suggests a direct link between 3D genome architecture and evolutionary divergence.

Intriguingly, we observed sex-specific differences in *D. herbacea*, the SDRs form distinct and highly structured chromatin domains, with this organization being more pronounced in the female SDR **(paper 3.4)**. This pattern is reminiscent of the *Marchantia* V (sex) chromosome, which is enriched in heterochromatic histone modification marks and TEs, and exhibits prominent self-interaction domains. Additionally, in *Marchantia*, gene expression correlates positively with chromatin domain boundaries, suggesting functional compartmentalization within the V chromosome[115]. These parallels indicate that it would be highly informative to investigate how the chromatin organization of the sex chromosome changes across the sexual cycle of brown algae in future studies.

Evolutionary analyses monoicous brown algae species *C. linearis* and *D. dudresnayi* have revealed the persistence of ancestral male genetic background of *Ectocarpus* sp.7 and *D. herbacea* [116]. In our comparisons between *D. herbacea* and closely related hermaphroditic species showed that sex-homologous chromosomes retain male-specific chromatin signatures even after the transition to hermaphroditism **(paper 3.4)**. In brown algae, hermaphrodites are thought to have arisen by acquired female-specific genes[116,117]

from ancestral male lineages. Consistent with this, our results uncover that V-specific chromatin architecture maintained through this transition.

Fucus species have XY sex chromosomes, which differ from the sexual systems of most brown algae (**paper 3.2**). It would therefore be interesting to investigate chromatin conformation changes associated with the transition to dioicy. In addition, sex chromosomes are typically enriched in transposable elements and heterochromatic regions; future studies integrating histone post-translational modification and Hi-C data could provide valuable insights into how chromatin states of sex chromosomes differ from those of autosomes.

# 5 Conclusion

The genome is folded hierarchically within the nucleus into CTs, compartment A/B and chromatin domains. However, the biological roles and evolutionary patterns of these structures are still poorly characterized, especially in non-model multicellular lineage as brown algae. By using Nanopore and Hi-C data, we significantly improved reference genomes for six brown algal species and generated new chromosome-level assemblies for two female strains. Coupled with transcriptomes, we comprehensively characterized the genome structures and dissected the principles underlying the 3D genome architecture of six brown algae and one outgroup species. Interaction frequency Hi-C maps we observed reveal a remarkable diversity in 3D genome organization, ranging from species with simple chromatin organization (associated with small genomes) to those exhibiting structured, TAD-like domains (associated with large, TE-enriched genomes).

Evolutionary analysis indicates that the overall chromatin architecture has been maintained for approximately 11 Myr. In-depth analysis surprisingly uncovered a deep conservation of sex chromosome 3D structure, persisting despite changes in sexual systems across the species. Finally, through tracing the evolutionary origins of brown algal centromeres, we revealed their remarkable, long-term co-evolution with centrophilic retrotransposons and identify them as persistent hotspots of chromosomal breakage and rearrangement.

In summary, this thesis moves beyond simply describing genome structure and provides strong evidence for how the evolutionary forces acting on TEs, centromeres, and sexual systems have fundamentally shaped the 3D architecture of the nucleus in a major eukaryotic lineage.

# Bibliography


1. Dekker, J. & Mirny, L. A. The chromosome folding problem and how cells solve it. *Cell* **187**, 6424–6450 (2024).
2. Paweletz, N. Walther Flemming: pioneer of mitosis research. *Nat. Rev. Mol. Cell Biol.* **2**, 72–75 (2001).
3. Cremer, T. *et al.* Rabl's model of the interphase chromosome arrangement tested in Chinise hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. *Hum. Genet.* **60**, 46–56 (1982).
4. Boveri, T. Die Blastomerenkerne von Ascaris megalocephala und die Theorie der Chromosomenindividualität. *Arch Zellforsch* **3**, 181–268 (1909).
5. Kornberg, R. D. Chromatin Structure: A Repeating Unit of Histones and DNA. *Science* **184**, 868–871 (1974).
6. Finch, J. T. & Klug, A. Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci.* **73**, 1897–1901 (1976).
7. Thoma, F., Koller, T. & Klug, A. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J. Cell Biol.* **83**, 403–427 (1979).
8. Borden, J. & Manuelidis, L. Movement of the X chromosome in epilepsy. *Science* **242**, 1687–1691 (1988).
9. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, (2009).
10. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).

11. Álvarez-González, L. & Ruiz-Herrera, A. Evolution of 3D Chromatin Folding. *Annu. Rev. Anim. Biosci.* **29**, 8 (2025).

12. Dong, P., Tu, X., Liang, Z., Kang, B. H. & Zhong, S. Plant and animal chromatin three-dimensional organization: Similar structures but different functions. *J. Exp. Bot.* **71**, 5119–5128 (2020).

13. Szalay, M. F., Majchrzycka, B., Jerković, I., Cavalli, G. & Ibrahim, D. M. Evolution and function of chromatin domains across the tree of life. *Nat. Struct. Mol. Biol.* **31**, 1824–1837 (2024).

14. Markaki, Y. *et al.* Xist nucleates local protein gradients to propagate silencing across the X chromosome. *Cell* **184**, 6174-6192.e32 (2021).

15. Hsieh, T.-H. S. *et al.* Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol. Cell* **78**, 539-553.e8 (2020).

16. Ing-Simmons, E. *et al.* Independence of chromatin conformation and gene regulation during Drosophila dorsoventral patterning. *Nat. Genet.* **53**, 487–499 (2021).

17. Han, J., Zhang, Z. & Wang, K. 3C and 3C-based techniques: The powerful tools for spatial genome organization deciphering. *Mol. Cytogenet.* **11**, 21 (2018).

18. Sun, L. *et al.* Mapping nucleosome-resolution chromatin organization and enhancer-promoter loops in plants using Micro-C-XL. *Nat. Commun.* **15**, 35 (2024).

19. Abbas, A. *et al.* Integrating Hi-C and FISH data for modeling of the 3D organization of chromosomes. *Nat. Commun.* **10**, 2049 (2019).

20. Lukyanchikova, V. *et al.* Anopheles mosquitoes reveal new principles of 3D genome organization in insects. *Nat. Commun.* **13**, 1960 (2022).

21. Bredeson, J. V. *et al.* Conserved chromatin and repetitive patterns reveal slow genome evolution in frogs. *Nat. Commun.* **15**, (2024).

22. Hoencamp, C. *et al.* 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. *Science* **372**, (2021).

23. Hildebrand, E. M. & Dekker, J. Mechanisms and Functions of Chromosome Compartmentalization. *Trends Biochem. Sci.* **45**, 385–396 (2020).

24. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

25. Contessoto, V. G. *et al.* Interphase chromosomes of the Aedes aegypti mosquito are liquid crystalline and can sense mechanical cues. *Nat. Commun.* **14**, 326 (2023).

26. Liu, C. *et al.* Genome-wide analysis of chromatin packing in Arabidopsis thaliana at single-gene resolution. *Genome Res.* **26**, 1057–1068 (2016).

27. Liao, Y., Zhang, X., Chakraborty, M. & Emerson, J. J. Topologically associating domains and their role in the evolution of genome structure and function in Drosophila. *Genome Res.* **31**, 397–410 (2021).

28. Kurbidaeva, A. *et al.* Topologically associating domains and the evolution of three-dimensional genome architecture in rice. *Plant J.* **122**, e70139 (2025).

29. Cremer, T. & Cremer, M. Chromosome Territories. *Cold Spring Harb. Perspect. Biol.* **2**, a003889 (2010).

30. Schneider, R. & Grosschedl, R. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes Dev.* **21**, 3027–3043 (2007).

31. Hildebrand, E. M. *et al.* Mitotic chromosomes are self-entangled and disentangle through a topoisomerase-II-dependent two-stage exit from mitosis. *Mol. Cell* **84**, 1422-1441.e14 (2024).

32. Liu, J. From Rabl-like Architecture to Chromosome Territories: A Conserved Developmental Transition in Animal Genomes. *Mol. Biol. Evol.* **42**, msaf235 (2025).

33. Tourdot, E. & Grob, S. Three-dimensional chromatin architecture in plants – General features and novelties. *Eur. J. Cell Biol.* **102**, 151344 (2023).

34. Zhang, F. *et al.* The F-Box Protein ZYGO1 Mediates Bouquet Formation to Promote Homologous Pairing, Synapsis, and Recombination in Rice Meiosis. *Plant Cell* **29**, 2597–2609 (2017).

35. Concia, L. *et al.* Wheat chromatin architecture is organized in genome territories and transcription factories. *Genome Biol.* **21**, 104 (2020).

36. Dong, P. *et al.* 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments. *Mol. Plant* **10**, 1497–1509 (2017).

37. Sun, J. *et al.* Genomic and epigenomic insight into giga-chromosome architecture and adaptive evolution of royal lily (Lilium regale). *Nat. Commun.* **16**, 5617 (2025).

38. Sun, L. *et al.* Heat stress-induced transposon activation correlates with 3D chromatin organization rearrangement in Arabidopsis. *Nat. Commun.* **11**, 1886 (2020).

39. Carlton, P. M. & Cande, W. Z. Telomeres act autonomously in maize to organize the meiotic bouquet from a semipolarized chromosome orientation. *J. Cell Biol.* **157**, 231–242 (2002).

40. Harris, H. L. *et al.* Chromatin alternates between A and B compartments at kilobase scale for subgenic organization. *Nat. Commun.* **14**, 3303 (2023).

41. Xiao, S., Luo, L., Yang, M., He, H. & Zhou, Y. Fine-scale 3D chromatin architectures and their regulatory mechanisms in plants. *Curr. Opin. Plant Biol.* **88**, 102786 (2025).

42. Grob, S., Schmid, M. W. & Grossniklaus, U. Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the flamenco Locus of Drosophila. *Mol. Cell* **55**, 678–693 (2014).

43. Concia, L. *et al.* Genome-Wide Analysis of the Arabidopsis Replication Timing Program. *Plant Physiol.* **176**, 2166–2185 (2018).

44. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).

45. Song, Z. *et al.* Dynamic changes in 3D chromatin structure during male gametogenesis in Arabidopsis thaliana. *Genome Biol.* **26**, 27 (2025).

46. Sandoval-Velasco, M. *et al.* Three-dimensional genome architecture persists in a 52,000-year-old woolly mammoth skin sample. *Cell* **187**, 3541-3562.e51 (2024).

47. Liao, Y. *et al.* The 3D architecture of the pepper genome and its relationship to function and evolution. *Nat. Commun.* **13**, (2022).

48. Falk, M. *et al.* Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570**, 395–399 (2019).

49. Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin Domains: The Unit of Chromosome Organization. *Mol. Cell* **62**, 668–680 (2016).

50. Bürmann, F. & Löwe, J. Structural biology of SMC complexes across the tree of life. *Curr. Opin. Struct. Biol.* **80**, 102598 (2023).

51. Lawson, H. A., Liang, Y. & Wang, T. Transposable elements in mammalian chromatin organization. *Nat. Rev. Genet.* **24**, 712–723 (2023).

52. Ali, T., Renkawitz, R. & Bartkuhn, M. Insulators and domains of gene expression. *Curr. Opin. Genet. Dev.* **37**, 17–26 (2016).

53. Kitamura, Y. *et al.* CTCF-mediated 3D chromatin sets up the gene expression program in the male germline. *Nat. Struct. Mol. Biol.* **32**, 1227–1240 (2025).

54. Kahn, T. G. *et al.* Topological screen identifies hundreds of Cp190- and CTCF-dependent Drosophila chromatin insulator elements. *Sci. Adv.* **9**, eade0090 (2023).

55. Shu, J. *et al.* EMF1 functions as a 3D chromatin modulator in Arabidopsis. *Mol. Cell* **84**, 4729-4739.e6 (2024).

56. Luo, L., Yang, M. & Zhou, Y. BMI1s interact with condensin complexes to regulate chromatin 3D structure and gene expression in Arabidopsis. *aBIOTECH* **6**, (2025).

57. Göbel, A. M. *et al.* Mutations of PDS5 genes enhance TAD-like domain formation Arabidopsis thaliana. *Nat. Commun.* **15**, (2024).

58. Zhang, X. *et al.* Microrchidia ATPases and DNA 6mA demethylase ALKBH1 act antagonistically on PRC2 to control chromatin structure and stress tolerance. *Nat. Plants* **11**, 1591–1607 (2025).

59. Corbo, M., Damas, J., Bursell, M. G. & Lewin, H. A. Conservation of chromatin conformation in carnivores. *Proc. Natl. Acad. Sci. U. S. A.* **119**, (2022).

60. Álvarez-González, L. *et al.* Principles of 3D chromosome folding and evolutionary genome reshuffling in mammals. *Cell Rep.* **41**, (2022).

61. Marín-Gual, L. *et al.* Meiotic cohesin RAD21L shapes 3D genome structure and transcription in the male germline. *Sci. Adv.* **11**, eadv2283 (2025).

62. Davidson, I. F. & Peters, J.-M. Genome folding through loop extrusion by SMC complexes. *Nat. Rev. Mol. Cell Biol.* **22**, 445–464 (2021).

63. Kim, I. V. Chromatin loops are an ancestral hallmark of the animal regulatory genome. *Nature* 1–9 (2025) doi:10.1038/s41586-025-08960-w.

64. Wright, C. J. *et al.* Constraints on chromosome evolution revealed by the 229 chromosome pairs of the Atlas blue butterfly. *Curr. Biol.* **0**, (2025).

65. Zhou, B. *et al.* Evolutionary patterns and functional effects of 3D chromatin structures in butterflies with extensive genome rearrangements. *Nat. Commun.* **15**, 6303 (2024).

66. Yang, H. *et al.* A map of cis-regulatory elements and 3D genome structures in zebrafish. *Nature* **588**, 337–343 (2020).

67. Karaaslan, E. S. *et al.* Marchantia TCP transcription factor activity correlates with three-dimensional chromatin structure. *Nat. Plants* **6**, 1250–1261 (2020).

68. James, C., Trevisan-Herraz, M., Juan, D. & Rico, D. Evolutionary analysis of gene ages across TADs associates chromatin topology with whole-genome duplications. *Cell Rep.* **43**, (2024).

69. Principles of 3D chromosome folding and evolutionary genome reshuffling in mammals.

70. Kakui, Y. *et al.* Chromosome arm length, and a species-specific determinant, define chromosome arm width. *Cell Rep.* **41**, 111753 (2022).

71. Lu, J. Y. *et al.* Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. *Cell Res.* **31**, 613–630 (2021).

72. Kruse, K. *et al.* Transposable elements drive reorganisation of 3D chromatin during early embryogenesis. 523712 Preprint at https://doi.org/10.1101/523712 (2019).

73. Torres, D. E. *et al.* Implications of the three-dimensional chromatin organization for genome evolution in a fungal plant pathogen. *Nat. Commun.* **15**, (2024).

74. Li, S.-M. H. *et al.* Skin regional specification and higher-order HoxC regulation. *Sci. Adv.* **11**, eado2223 (2025).

75. Zhao, Y. *et al.* A compendium and comparative epigenomics analysis of cis-regulatory elements in the pig genome. *Nat. Commun.* **12**, 2217 (2021).

76. Mohan, A. V., Escuer, P., Cornet, C. & Lucek, K. A three-dimensional genomics view for speciation research. *Trends Genet.* **40**, 638–641 (2024).

77. Cock, J. M. *et al.* The Ectocarpus genome and the independent evolution of multicellularity in brown algae. *Nature* **465**, 617–621 (2010).

78. Coelho, S. M., Peters, A. F., Müller, D. & Cock, J. M. Ectocarpus: An evo-devo model for the brown algae. *EvoDevo* **11**, (2020).

79. Bringloe, T. T. *et al.* Phylogeny and Evolution of the Brown Algae. *Crit. Rev. Plant Sci.* **39**, 281–321 (2020).

80. Denoeud, F. *et al.* Evolutionary genomics of the emergence of brown algae as key components of coastal ecosystems. *Cell* **187**, 6943-6965.e39 (2024).

81. Coelho, S. M., Gueno, J., Lipinska, A. P., Cock, J. M. & Umen, J. G. UV Chromosomes and Haploid Sexual Systems. *Trends Plant Sci.* **23**, 794–807 (2018).

82. Bourdareau, S. *et al.* Histone modifications during the life cycle of the brown alga Ectocarpus. *Genome Biol.* **22**, 12 (2021).

83. Arun, A. *et al.* Convergent recruitment of TALE homeodomain life cycle regulators to direct sporophyte development in land plants and brown algae. *eLife* **8**, (2019).

84. Cossard, G. G. *et al.* Selection drives convergent gene expression changes during transitions to co-sexuality in haploid sexual systems. *Nat. Ecol. Evol.* **6**, (2022).

85. Barrera-Redondo, J. *et al.* Origin and evolutionary trajectories of brown algal sex chromosomes. *Nat. Ecol. Evol.* 1–18 (2025) doi:10.1038/s41559-025-02838-w.

86. Lotharukpong, J. S. *et al.* A transcriptomic hourglass in brown algae. *Nature* https://doi.org/10.1038/s41586-024-08059-8 (2024) doi:10.1038/s41586-024-08059-8.

87. Heesch, S. *et al.* Evolution of life cycles and reproductive traits: Insights from the brown algae. *J. Evol. Biol.* **34**, 992–1009 (2021).

88. Zhu, Z., Younas, L. & Zhou, Q. Evolution and regulation of animal sex chromosomes. *Nat. Rev. Genet.* **26**, 59–74 (2024).

89. Sahara, K., Yoshido, A. & Traut, W. Sex chromosome evolution in moths and butterflies. *Chromosome Res.* **20**, 83–94 (2012).

90. Bachtrog, D. *et al.* Are all sex chromosomes created equal? *Trends Genet.* **27**, 350–357 (2011).

91. Ahmed, S. *et al.* A haploid system of sex determination in the brown alga ectocarpus sp. *Curr. Biol.* **24**, 1945–1957 (2014).

92. Mignerot, L. & Coelho, S. M. The origin and evolution of the sexes: Novel insights from a distant eukaryotic linage. *C. R. Biol.* **339**, 252–257 (2016).

93. Choi, S. W. *et al.* Ordovician origin and subsequent diversification of the brown algae. *Curr. Biol.* **34**, 740-754.e4 (2024).

94. Barrera-Redondo, J. *et al.* Origin and evolutionary trajectories of brown algal sex chromosomes. *Nat. Ecol. Evol.* 1–18 (2025) doi:10.1038/s41559-025-02838-w.

95. Denoeud, F. *et al.* Evolutionary genomics of the emergence of brown algae as key components of coastal ecosystems. *Cell* **187**, 6943-6965.e39 (2024).

96. Bracewell, R., Chatla, K., Nalley, M. J. & Bachtrog, D. Dynamic turnover of centromeres drives karyotype evolution in Drosophila. *eLife* **8**, e49002 (2019).

97. Sankaranarayanan, S. R. *et al.* Loss of centromere function drives karyotype evolution in closely related Malassezia species. *eLife* **9**, e53944 (2020).

98. Huang, Z. *et al.* Recurrent chromosome reshuffling and the evolution of neo-sex chromosomes in parrots. *Nat. Commun.* **13**, (2022).

99. Immler, S. & Otto, S. P. The evolution of sex chromosomes in organisms with separate haploid sexes. *Evol. Int. J. Org. Evol.* **69**, 694–708 (2015).

100. Liu, P. *et al.* 3D chromatin maps of a brown alga reveal U/V sex chromosome spatial organization. *Nat. Commun.* **15**, 9590 (2024).

101. Harmston, N. *et al.* Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nat. Commun.* **8**, 441 (2017).

102. Dinatale, E., Craig, R. J., Martinho, C., Drost, H.-G. & Coelho, S. M. Characterization of the transposable element landscape shaping the Ectocarpus genome. *Genome Biol.* **26**, 1–33 (2025).

103. Naish, M. & Henderson, I. R. The structure, function, and evolution of plant centromeres. *Genome Res.* **34**, 161–178 (2024).

104. Bousios, A., Kakutani, T. & Henderson, I. R. Centrophilic Retrotransposons of Plant Genomes. *Annu. Rev. Plant Biol.* **76**, 579–604 (2025).

105. Henikoff, S., Ahmad, K. & Malik, H. S. The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA. *Science* **293**, 1098–1102 (2001).

106. Tsukahara, S. *et al.* Centrophilic retrotransposon integration via CENH3 chromatin in Arabidopsis. *Nature* **637**, 744–748 (2025).

107. Gao, X., Hou, Y., Ebina, H., Levin, H. L. & Voytas, D. F. Chromodomains direct integration of retrotransposons to heterochromatin. *Genome Res.* **18**, 359–369 (2008).

108. Sharma, A. & Presting, G. G. Centromeric retrotransposon lineages predate the maize/rice divergence and differ in abundance and activity. *Mol. Genet. Genomics MGG* **279**, 133–147 (2008).

109. Houben, A. *et al.* CENH3 interacts with the centromeric retrotransposon cereba and GC-rich satellites and locates to centromeric substructures in barley. *Chromosoma* **116**, 275–283 (2007).

110. Wolfgruber, T. K. *et al.* Maize Centromere Structure and Evolution: Sequence Analysis of Centromeres 2 and 5 Reveals Dynamic Loci Shaped Primarily by Retrotransposons. *PLoS Genet.* **5**, e1000743 (2009).

111. Helsen, J., Ramachandran, K., Sherlock, G. J. & Dey, G. Centromeres evolve progressively through selection at the kinetochore interface. (2025) doi:10.1101/2025.01.16.633479.

112. Sun, X., Wahlstrom, J. & Karpen, G. Molecular Structure of a Functional Drosophila Centromere. *Cell* **91**, 1007–1019 (1997).

113. Neumann, P. *et al.* Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mob. DNA* **2**, 4 (2011).

114. Ostromyshenskii, D. I., Chernyaeva, E. N., Kuznetsova, I. S. & Podgornaya, O. I. Mouse chromocenters DNA content: sequencing and in silico analysis. *BMC Genomics* **19**, 151 (2018).

115. Montgomery, S. A. *et al.* Chromatin Organization in Early Land Plants Reveals an Ancestral Association between H3K27me3, Transposons, and Constitutive Heterochromatin. *Curr. Biol.* **30**, 573-588.e7 (2020).

116. Barrera-Redondo, J. *et al.* Origin and evolutionary trajectories of brown algal sex chromosomes. *Nat. Ecol. Evol.* https://doi.org/10.1038/s41559-025-02838-w (2025) doi:10.1038/s41559-025-02838-w.

117. Cossard, G. G. *et al.* Selection drives convergent gene expression changes during transitions to co-sexuality in haploid sexual systems. *Nat. Ecol. Evol.* **6**, 579–589 (2022).

# Appendix

nature communications

Article

https://doi.org/10.1038/s41467-024-53453-5

# 3D chromatin maps of a brown alga reveal U/V sex chromosome spatial organization

Received: 23 April 2024

Accepted: 8 October 2024

Published online: 06 November 2024

Check for updates

Pengfei Liu[1], Jeromine Vigneau[1], Rory J. Craig[1], Josué Barrera-Redondo[1], Elena Avdievich[1], Claudia Martinho[1,3], Michael Borg[1], Fabian B. Haas[1], Chang Liu[2] & Susana M. Coelho[1]

Nuclear three dimensional (3D) folding of chromatin structure has been linked to gene expression regulation and correct developmental programs, but little is known about the 3D architecture of sex chromosomes within the nucleus, and how that impacts their role in sex determination. Here, we determine the sex-specific 3D organization of the model brown alga *Ectocarpus* chromosomes at 2 kb resolution, by mapping long-range chromosomal interactions using Hi-C coupled with Oxford Nanopore long reads. We report that *Ectocarpus* interphase chromatin exhibits a non-Rabl conformation, with strong contacts among telomeres and among centromeres, which feature centromere-specific LTR retrotransposons. The *Ectocarpus* chromosomes do not contain large local interactive domains that resemble TADs described in animals, but their 3D genome organization is largely shaped by post-translational modifications of histone proteins. We show that the sex determining region (SDR) within the U and V chromosomes are insulated and span the centromeres and we link sex-specific chromatin dynamics and gene expression levels to the 3D chromatin structure of the U and V chromosomes. Finally, we uncover the unique conformation of a large genomic region on chromosome 6 harboring an endogenous viral element, providing insights regarding the impact of a latent giant dsDNA virus on the host genome's 3D chromosomal folding.

Sex chromosomes are unique genomic regions that evolved independently many times in different groups of eukaryotes. Three types of sex chromosome systems exist in nature, the well-described XX/XY and ZW/ZZ systems and the still elusive U/V systems, in organisms that express sex in the haploid stage of the life cycle[1–6]. Heteromorphic sex chromosomes (Y, W, U, V) have evolved repeatedly in diverse eukaryotic species. Suppression of recombination between X and Y (or Z and W, or U and V) chromosomes usually leads to a range of genomic modifications in these regions, including degeneration of the non-recombining chromosome, accumulation of repeats, and gene decay via accumulation of deleterious mutations[7,8]. Repeats pose the largest challenge for reference genome assembly, and centromeres, subtelomeres, and repeat-rich sex chromosomes are typically ignored in sequencing projects. Consequently, complete sequence assemblies of heteromorphic Y, W, U, and V sex chromosomes have only been generated across a handful of taxa[9–16], and most of the information is fragmentary even at the linear sequence level. Moreover, despite the key association between the three-dimensional (3D) structure of chromatin and gene regulation[17,18], we lack critical information regarding chromatin landscapes and nuclear 3D organization of sex chromosomes

[1]Department of Algal Development and Evolution, Max Planck Institute for Biology Tübingen, Tübingen, Germany. [2]Institute of Biology, University of Hohenheim, Stuttgart, Germany. [3]Present address: School of Life Sciences, Division of Plant Sciences, University of Dundee, At James Hutton Institute, Errol Road, Invergowrie, Dundee, UK. e-mail: susana.coelho@tuebingen.mpg.de

within the nuclear space, and how chromatin folding is associated with the sex-specific gene expression underlying sexual differentiation.

Genome folding generally involves hierarchical structures ranging from chromatin loops to chromosome territories[19]. The best-known 3D chromatin organization units are topologically associating domains (TADs), which show a self-interacting pattern with strongly interacting boundaries in Hi-C contact maps of animal genomes[20]. Genome architectural proteins, such as CTCF (CCCTC-binding factor) and cohesin, bind strongly to DNA anchor sites and mediate the formation of chromatin contact domains through loop extrusion[21]. In addition to TADs, structural units called compartmental domains have been demonstrated in animals (e.g.,[22]). Compartmental domains are closely associated with local chromatin states and preferentially interact with other compartmental domains of similar chromatin states, contributing to the establishment of the 3D architecture for a given genome[19,22]. Plant genomes also frequently exhibit a higher-order 3D chromatin organization. TADs or TAD-like structures have been described in several plant species[23], although their genomes do not encode CTCF homologs[24]. In contrast to animal TADs that have sharp and well-delineated boundaries on Hi-C maps, plant TADs exhibit less pronounced boundaries due to weaker chromatin insulation[25]. In contrast, *Arabidopsis* (*Arabidopsis thaliana*) lacks plant TADs[26]. The absence of TADs in the *Arabidopsis* genome is likely related to its small size, high gene density, and short intergenic regions. However, chromatin loops and A/B compartments are present in *Arabidopsis* (e.g.,[27,28]), and small structural units within 3D chromatin architecture have been recently described[29].

Here, we generated 2 kb-resolution maps of the male and female haploid genomes of the brown algal model *Ectocarpus* and examined the 3D chromatin structure of autosomes compared to U and V sex chromosomes. The *Ectocarpus* life cycle involves an alternation between diploid and haploid generations, with sex being determined in the haploid stage of the life cycle by U (female) and V (male) sex chromosomes[5]. Therefore, this model organism provides the opportunity to investigate the U/V sex chromosome organization in comparison to autosomes. Our near-complete assembly of the *Ectocarpus* genome (*Ectocarpus* V5) offers an improved reference genome and allows us to define and characterize the centromeric and subtelomeric sequences in this organism. We found that interphase chromatin is organized in a non-Rabl configuration, with telomeres and centromeres of all 27 *Ectocarpus* chromosomes clustering together in the 3D nuclear space. We reveal that the 3D structure of *Ectocarpus* chromatin is highly streamlined, not organized into TADs, and A and B chromatin compartments are mainly defined by H3K79me2 deposition and depletion of activation marks. We then focus on the 3D structure of the *Ectocarpus* U and V sex chromosomes to show that in both sex chromosomes, the SDR spans the centromere, and is insulated from the rest of the chromosome. We found no overall differences in the 3D chromatin organization between male (V) and female (U) chromosomes but both have different 3D organization compared with autosomes. Finally, we uncover the distinctive conformation of a genomic region on chromosome 6 harboring a giant endogenous viral element (EVE), giving insights into the interplay of dsDNA viruses with the chromatin environment in the host.

## Results

### A near complete assembly of the male and female haploid genome of *Ectocarpus*

A complete assembly of the *Ectocarpus* genome has been challenging mainly due to the presence of highly repetitive regions, which short-read Illumina sequencing, low coverage Hi-C, and Sanger sequencing could not hitherto successfully resolve. The published version of the *Ectocarpus sp. 7* reference genome (V2) contains 28 pseudo-chromosomes spanning 176.99 Mb, with 17.97 Mb of unplaced contigs, a contig-level N50 of 33 kb, and a total of 11,588 gaps[30]. Here, we combined Oxford Nanopore Technologies (ONT) long reads and Hi-C sequencing techniques to achieve near-complete assemblies of both the haploid male and female genome of *Ectocarpus sp. 7* (Fig. 1A and Supplementary Data 1, 2).

The ONT long reads were obtained separately from male and female siblings (Ec32m, Ec25f, Supplementary Figs. 1, 2), totaling 11 Gb and 20 Gb of data, respectively. ONT long-reads were complemented with Hi-C data, encompassing 822 million pairs of sequences (~ 135 Gb) at a sequencing depth of 635x for the male, and ~ 444 million pairs (~ 73 Gb) and 338x coverage for the female (see "Methods" for details) (Supplementary Fig. 1 and Supplementary Data 3). The *Ectocarpus* V5 male genome assembly has an N50 of 7.0 Mb and a total size of 186.6 Mb. Chromosome 28 from V2 is now part of chromosome 4, bringing the total number of chromosomes to 27, with sizes ranging from 4.52 Mb to 10.73 Mb. Similarly to what was done for the *Ectocarpus* V2[30], we added the female SDR (size 1.55 Mb) to this male genome in order to obtain a final *Ectocarpus* V5 reference genome (Supplementary Fig. 1).

The genome is highly contiguous: out of the 27 assembled chromosome models, most are gapless, and only six chromosomes have 10 gaps in total (Fig. 1A and Supplementary Data 4). The accuracy of Hi-C-based chromosome construction was evaluated manually by inspecting the chromatin contact matrix at 100 kb resolution, which exhibited a well-organized interaction contact pattern along the diagonals within each pseudo-chromosome (Fig. 1B).

Telomeric regions were almost entirely absent from earlier genome assemblies, although a putative telomere bearing the repeat (TTAGGG)n was observed[31]. In our *Ectocarpus* V5 assembly, 43 of the 54 telomeric regions are fully resolved, and twelve of the 27 pseudo-chromosomes correspond to a telomere-to-telomere assembly (Fig. 1A and Supplementary Fig. 3). On all but three of the resolved telomeric regions, we observed a specific satellite repeat adjacent to the telomeric repeats. The satellite features a repeated monomer of ~ 98 bp and is almost exclusively found at the sub-telomeres, where it forms arrays that range from only a few to more than 100 copies. Notably, the telomeric motif TTAGGG is present in three independent locations within each satellite monomer (Supplementary Fig. 4). Similar subtelomeric organizations have been observed in several species, including the green alga *Chlamydomonas reinhardtii*, where the telomere-like motifs present within the sub-telomeric satellites are hypothesized to serve as seed sequences that facilitate telomere healing following DNA damage[32]. Eight of the V5 chromosome arms terminated in the subtelomeric repeat, leaving only four chromosome extremities for which the assembly failed to reach either the subtelomere or telomere (Fig. 1A).

Ribosomal DNA (rDNA) arrays were also poorly resolved in previous assemblies. The V5 assembly revealed a single major rDNA array located within an internal region of chromosome 4 (position: 269–274 Mb), which features six rDNA repeats before collapsing the assembly due to high levels of repeat content. The 5S rDNA gene is linked to the main rDNA unit (18S-5.8S-26S), as previously reported in many brown algae and Stramenopiles[33]. Considering an ONT read coverage fold change of approximately 20 between the rDNA array and the flanking regions, we estimate that the rDNA array may consist of > 100 rDNA repeats, spanning ~ 1 Mb.

The total repeat content of the assembled chromosomes is estimated to be 29.8% (Supplementary Data 4). 13.75 Mb of additional sequence could not be assembled into chromosomes and remains unplaced in the V5 assembly. These sequences are highly repetitive (74.3% repeats) and presumably include heterochromatic regions that correspond to some of the assembly gaps or incomplete chromosome ends. Longer reads or alternative technologies will be required to achieve complete assembly of these complex regions.

https://doi.org/10.1038/s41467-024-53453-5

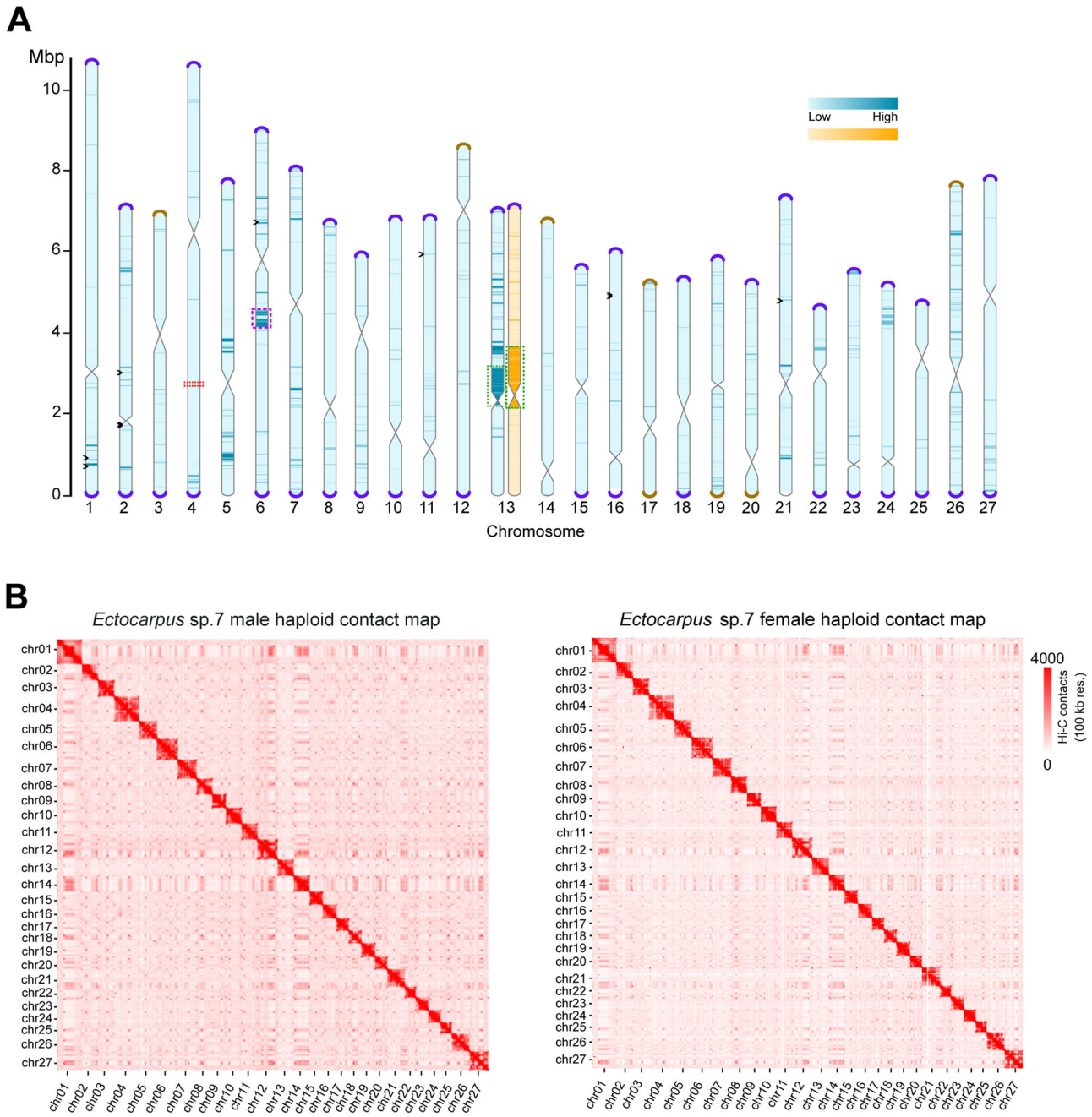



**Fig. 1 | *Ectocarpus sp. 7* whole-genome assembly. A** Schematic representation of the near telomere-to-telomere assembly of the 27 *Ectocarpus sp. 7* chromosomes, in haploid male (blue) and female (orange). Telomeres are represented as violet caps, sub-telomeres in brown. Centromeric regions are represented by the constrictions in the center of the chromosomes. The chromosomes are filled by variant density between the male and female haploid genomes used for the assembly (darker color means more differences). Violet dotted boxes represent the genomic region where a dsDNA virus is inserted, green dotted boxes represent the SDRs, and the red dotted box shows the rDNA array. Black arrowheads depict gaps. See methods for details. **B** Normalized genome-wide Hi-C contact map showing frequencies of pairwise 3D genome contacts at a 100 kb resolution in the male and female haploid genomes. The stripes seen in the contact map indicate regions of high contact frequency, corresponding to A/B compartments where regions within the same compartments interact more frequently with each other than with regions in other compartments. The dots scattered across the contact map represent specific loci that have higher contacts than the rest of the genome, often suggesting interactions between telomeres and centromeres or loops between these regions.

Since the V2 genome had a high-quality gene annotation, we performed a liftover of the V2 gene models to the *Ectocarpus* V5 genome. Out of the 18,412 V2 gene models, 18,278 could be lifted, and the remaining were mostly located on an unassigned scaffold in the V2 assembly. Genome completeness was quantified by BUSCO[34]. Two database sets were used, Eukaryota (255 core genes) and stramenopiles (100 core genes). Of the 255 core eukaryotic BUSCO genes, the V5 reference assembly contains 226 (88.7%) complete BUSCO genes. This represents a gain of 8 genes (+3.2%) compared to the V2 genome. The stramenopiles dataset resulted in 93% BUSCO completeness (increased by 1%; Supplementary Data 5).

## The *Ectocarpus* 3D chromatin architecture

To explore the 3D chromatin architecture of *Ectocarpus*, we mapped male and female Hi-C reads back to the V5 assembly (Supplementary Fig. 1 and Supplementary Data 3). Biological replicates were highly

https://doi.org/10.1038/s41467-024-53453-5

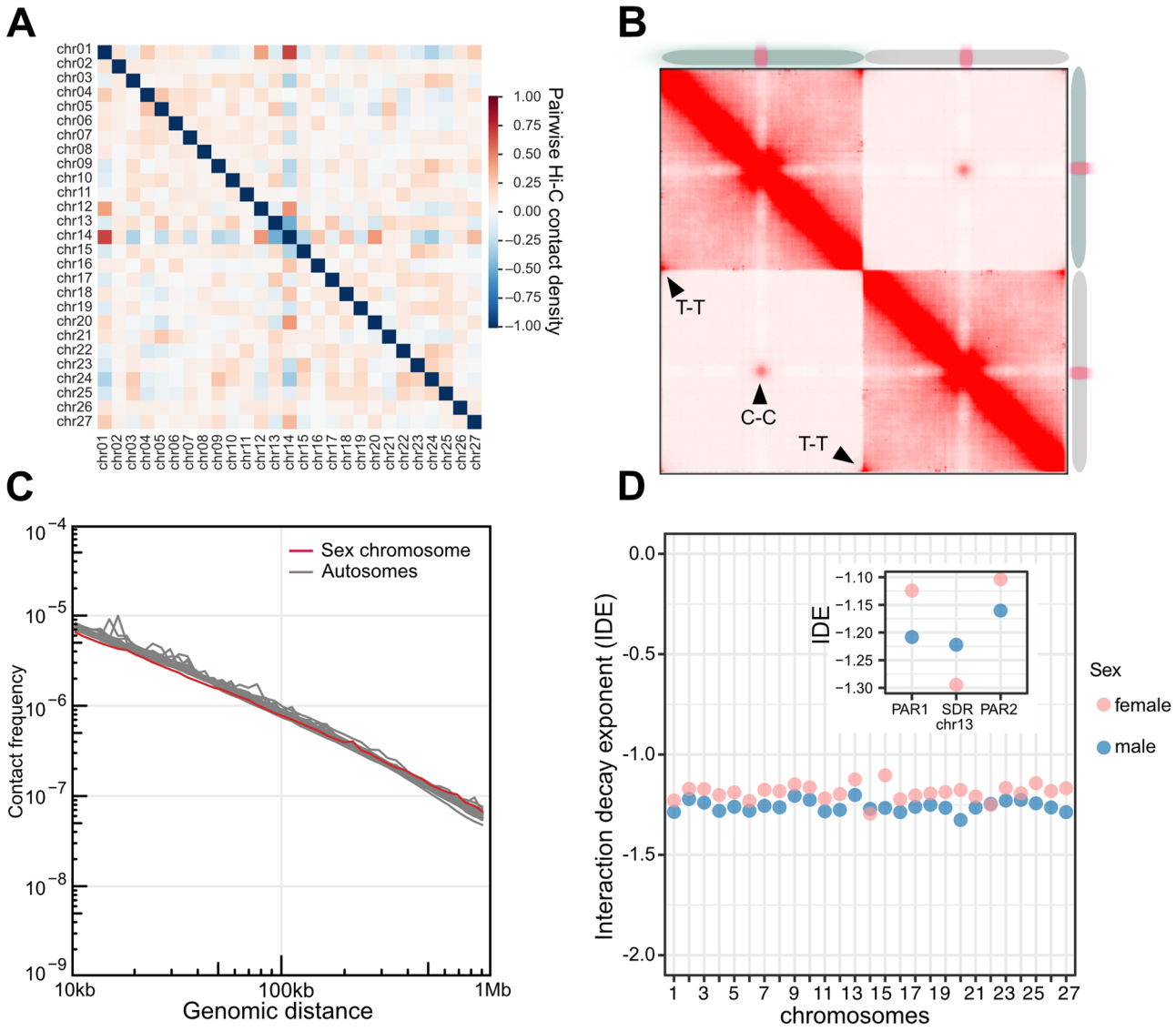



**Fig. 2 | 3D chromatin architecture of *Ectocarpus* revealed by Hi-C data. A** Pairwise averaged log-transformed observed/expected inter-chromosome contacts of *Ectocarpus* male at 10 k resolution. This scale represents the relative density of interactions between different chromosomes, with darker colors representing higher contact frequencies compared to what is expected by chance. **B** Analysis of aggregated intra- and inter-chromosomal contacts (Aggregate Chromosomal Analysis, ACA[104]), where individual chromosomes are linearly transformed to have the same length, and the centromere is placed at the center. The adjusted chromosomes are subsequently used to compute average intra- and inter-chromosomal contacts. The analysis shows how centromeres and telomeres of these chromosomes interact both within themselves and with each other. T-T: telomere to telomere interactions; C-C: centromere to centromere interactions. **C** Global folding patterns of each of the male *Ectocarpus* chromosomes reflected by contact frequency as a function of genomic distance (Ps). **D** IDEs of each autosome and sex chromosome region in *Ectocarpus* male and female. Normalized Hi-C matrices at a resolution of 10 kb at a distance range of 10 kb to 500 kb were used to calculate IDEs.

correlated (Pearson $r = 0.96$ and $r = 0.94$ for male and female samples, respectively, Supplementary Fig. 5), therefore, replicates were combined for downstream analysis to produce sex-specific high-resolution maps. We obtained 188.8 and 134.8 million interaction read pairs for male and female *Ectocarpus*, respectively, reaching a 2 kb resolution for each of the sexes.

In animals and plants, chromosomes are hierarchically packed in the nuclear space, and each occupies discrete regions referred to as a chromosome territory (CT)[23,35]. Chromosomal territories were detected in *Ectocarpus*, reflected by strong intra-chromosomal interactions and clear boundaries between chromosomes (see Fig. 1B). The global Hi-C maps of male and female *Ectocarpus* show no noticeable differences among autosomes, suggesting that the overall chromosomal territory organization is highly similar between the sexes. In addition, both the male and female sex chromosomes do not display any distinct intra- or inter-chromosomal contact patterns that differentiate them from autosomes. Therefore, the *Ectocarpus* genome folding on a broad chromosomal level appears to be consistent across both sexes and all chromosomes. We found a significant enrichment of inter-chromosomal interactions involving chromosomes 1, 12, 14, 20, and 27 (Fig. 2A), suggesting a propensity for these chromosomes to establish stronger contacts compared to others. Furthermore, strong contacts among telomeric regions of different chromosomes, as well as contacts among centromeric regions (see below), were widespread on the Hi-C map (Fig. 2B).

Next, we computed each chromosome's chromatin contact probability as a function of genomic distance to examine *Ectocarpus* chromosome packing patterns. As expected, we observed a decline in contact frequencies as genomic distances increased (Fig. 2C). Next, Interaction Decay Exponents (IDEs), which describe how fast

interaction frequencies drop with increasing physical genomic distance, were computed to characterize chromatin packaging[36,37]. We found that for each of the *Ectocarpus* chromosomes, interaction frequencies decayed in similar power-law functions with IDE values between 10 kb and 500 Kb (Fig. 2D). However, the IDE values in SDRs and PARs of sex chromosomes showed noticeable variation, suggesting differences in local chromatin packing in these regions (see below).

One prominent feature of animal and plant genomes is the organization of chromatin into TADs, characterized by preferential contacts between loci inside the same TAD and strong insulation from loci in adjacent TADs[20,38]. In some cases, TADs can promote enhancer-promoter contacts important for gene expression[39]. Intriguingly, we did not observe conspicuous TADs patterns in any of the *Ectocarpus* chromosomes upon zooming into the Hi-C map (Supplementary Fig. 6). Note that *Ectocarpus* has a similar genome size to the land plant *Arabidopsis*, which also does not exhibit classical TAD structure, but rather TAD-like domains that are moderately insulated from flanking chromatin regions[26,40] and are considered as an outlier species concerning plant TAD formation[41].

## A/B compartment dynamics in males versus females

Spatially distinct nuclear compartments are a prominent feature of 3D chromatin organization in eukaryotes[36]. A/B compartments, which generally correlate to active and repressed chromatin, respectively, can be identified with the first eigenvector (EV1) generated from the principal components analysis of the correlation heatmap (PCA)[36]. We applied PCA to individual chromosome's Hi-C maps normalized at 10 kb bin size to identify the two spatial compartments (Fig. 3A). The compartment that displayed stronger inter-chromosomal chromatin contacts was called 'A', whereas the 'B' compartment had lower inter-chromosomal contacts (Fig. 3B). Interestingly, the genomic regions bearing the centromeres corresponded to the A compartment (Fig. 3A, B). Further PCA analysis on the A compartment indicated that centromeres formed distinct sub-compartments, which were spatially separated from the rest of the A compartment regions (Supplementary Fig. 7).

Although different chromosomes had different proportions of compartment A and B, we noticed that the U and V sex chromosomes (chromosome 13) exhibit large stretches of regions associated to B compartment (Fig. 3B), suggesting they have a distinct overall configuration compared to autosomes (see below).

The *Ectocarpus* genome has been reported to have various histone post-transcriptional modifications (PTMs) associated with gene transcriptional activities[42,43]. We therefore asked whether chromatin associated with different A/B compartments exhibited different histone modification profiles. To this end, we used published epigenomic datasets for a range of histone PTMs from the same strains (Ec560, Ec561)[44] and mapped the ChIP-seq datasets to our V5 genome. We then examined the enrichment of each histone PTM specifically associated with genes in each A or B compartment (see "Methods"). We found that for both male and female genomes, the histone PTMs associated with active gene expression, such as H3K4me3, H3K9ac, H3K27ac and H3K36me3, were significantly enriched in the A compartment (Fig. 3C) although more modestly so for H3K36me3. Conversely, peaks of H3K79me2, a histone mark associated with repressed chromatin in *Ectocarpus*[42,43], were only marginally enriched in the B compartment albeit significantly (Fig. 3C). Furthermore, genes located within A compartment regions exhibited higher expression levels than those in the B compartments (Fig. 3D). Note that canonical repressive (heterochromatin) marks such as H3K9me2 and H3K27me3 are absent in *Ectocarpus*[42,43], with higher transcript abundance correlating more strongly with the gradual acquisition of activation-associated marks (H3K9ac, H3K27ac, H3K4me3 and H3K36me3[43]. Our observations thus suggest that transcriptional states appear to be the main driver of chromatin organization patterns in *Ectocarpus*, unlike what has been reported in other multicellular eukaryotes where the A/B compartments are enriched with euchromatic and heterochromatic chromatin, respectively[23,36]. Alternatively, *Ectocarpus* might harbor other yet-to-be-characterized repressive marks that could influence the organization of A/B compartments.

The A/B compartment assignment of the male and female *Ectocarpus* genomes was highly similar; nonetheless, 5.3% of the *Ectocarpus* chromatin exhibited different A/B compartment identities in male and female Hi-C maps (Fig. 3E). In animals, compartment status and boundaries may change during cell differentiation and correlate with changes in gene expression profiles[45]. We therefore investigated whether such changes in compartment annotation were associated with the expression patterns of sex-biased genes (SBG), i.e., genes that show a significant change in expression in males versus females[46,47]. We used RNA-seq datasets[44] and identified 2069 SBGs (see "Methods" for details, Supplementary Data 7). Depending on the expression preference, these SBGs were further annotated as male- and female-biased genes (MBGs and FBGs), respectively. SBGs were, however, not enriched in the regions where A/B compartment identity changed (Supplementary Data 8, Chi-square test $p = 0.0649$). MBGs in males were upregulated when in compartment A compared to FBGs, and the opposite was true in females (Fig. 3F), and we noticed that whilst MBGs in males still show greater expression associated to the A compartment, the overall expression levels were higher, regardless of the compartment. Therefore, the patterns of expression of SBGs were correlated with their association with histone PTMs and to the specific 3D chromatin organization in males versus females.

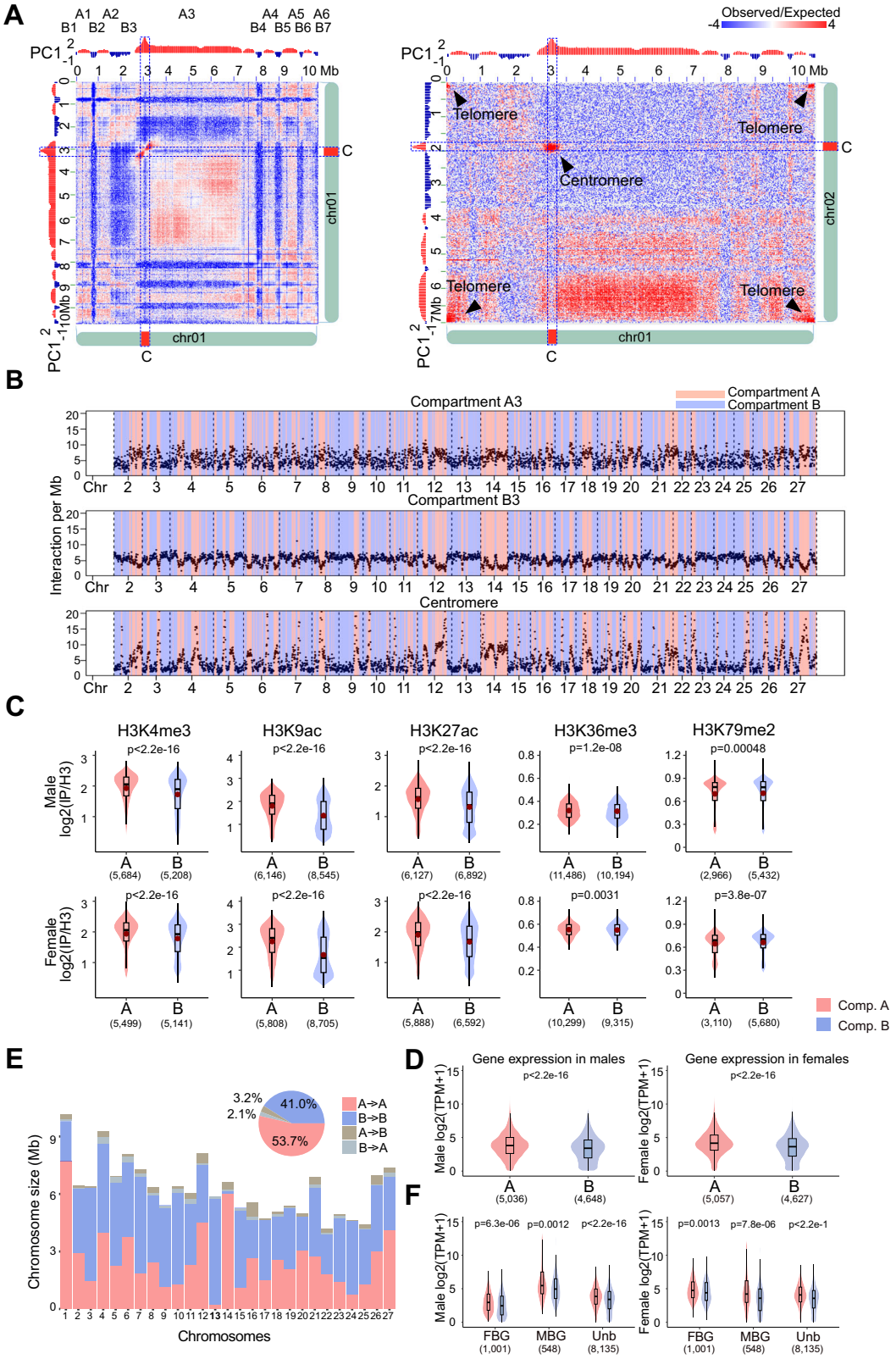
## U and V sex chromosomes and autosomes adopt distinct conformations

The sex-specific high-resolution genomic maps were then used to compare the sub-nuclear 3D genomic architecture of the U and V sex chromosomes. The U and V sex-specific regions (SDR) have been identified and characterized previously[9,48,49], but their largely repeat-rich nature has prevented their full assembly. In the *Ectocarpus* V5, the V and U chromosomes had a total length of 7.16 Mb and 7.23 Mb respectively (see Fig. 1). In *Ectocarpus*, U and V are largely homomorphic with a small region that is non-recombining (SDR) and therefore largely divergent between male and female[9,49] (Supplementary Fig. 8). The male and female SDRs of the *Ectocarpus* V5 genome feature no gaps. We also noticed that compared to the V2, the female SDR has increased in physical size. This was mainly due to the addition of repeats in the new assembly (V2 had 34.7% of repeats and V5 68.3% of repeats in the U-SDR). The small SDRs are flanked by large pseudoautosomal regions (PARs), which recombine at meiosis[9,50]. Structural analysis using our new assembly confirmed that the U and V sex chromosomes display unique characteristics compared with autosomes, including lower GC content, higher repeat content, lower gene density[9,50], and a largely repressive chromatin landscape[43] (Fig. 4A and Supplementary Data 4). We then used the 2 kb resolution Hi-C map to investigate the 3D structure of the sex chromosomes in the *Ectocarpus* nucleus. Intriguingly, the U and V sex chromosomes exhibited a distinct 3D architecture compared to autosomes, with their central SDRs both being insulated from the flanking PAR regions (Fig. 4B, see also Fig. 2D), with high intra-chromosomal contacts in the 3D space. We also noticed that both U and V SDRs spanned the centromeres (Fig. 4B).

## *Ectocarpus* centromeres are distinguished by specific LTR retrotransposons

To determine the structure and precise locations of the *Ectocarpus* centromeres, we analyzed the sequence characteristics of the chromosomal regions delineated by centromere-to-centromere interactions (see Figs. 2B, 3A, B). Regional centromeres vary extensively among eukaryotes, with common structures including short non-

https://doi.org/10.1038/s41467-024-53453-5




repetitive AT-rich regions, transposon-rich regions spanning tens to hundreds of kilobases, and megabase-scale satellite arrays[51]. We first searched for any specific repeat families that were (i) enriched in the putative centromeric regions, and (ii) common to all chromosomes. This revealed two retrotransposon families that are almost exclusively restricted to a single highly localized, gene-poor, and repeat-rich region on each chromosome (Fig. 5A and Supplementary Fig. 9). The most abundant of the two elements is a 6.6 kb *Metaviridae* (i.e., *Ty3/Gypsy*) long terminal repeat (LTR) retrotransposon, which encodes Gag and Pol on a single open reading frame of 1699 aa and can be found as full-length copies flanked by 4 bp target site duplications (Fig. 5B). The second is presumably a related LTR element, although it is only present in degraded fragments, and we were unable to recover an internal protein-coding region. The two retrotransposon families

https://doi.org/10.1038/s41467-024-53453-5

**Fig. 3 | High-resolution contact probability map reveals the higher-order organization of the *Ectocarpus* genome. A** Compartment A/B annotation based on principal component analysis. PC1 stands for the first principal component. The right panel shows inter-chromosomal contact patterns of A/B compartment regions between chromosomes 1 and 2. **B** Inter-chromosomal contacts of selected chromosome 1 regions with other chromosomes. The plots describe inter-chromosomal contacts belonging to the compartments A3 (top), B3 (middle), and centromere regions (bottom). The A/B compartment annotation of individual chromosomes is indicated with different colors. **C** Comparison of histone modifications, represented as log2(IP/H3) of regions enriched with selected histone marks. For each histone mark, the enriched regions are grouped according to the A/B compartments of *Ectocarpus* male and female. The mean value of log2(IP/H3) is represented by a red dot in each boxplot. Numbers in brackets represent the number of peaks of the corresponding histone ChIP-seq data. **D** Levels of gene expression in compartments (**A**) and (**B**) in males and females, represented as logs2(Transcripts Per Million + 1). *p*-values represent Wilcoxon tests. Numbers in brackets represent the number of genes. **E** Lengths of conserved and switching A/B compartment regions in male and female Ectocarpus genomes. "X - > Y" indicates compartment annotation in males ("X") and females ("Y"). The pie chart indicates pooled data from all chromosomes. **F** Expression of sex-biased genes (SBG) in compartment A/B regions. MBG: male-biased gene; FBG: female-biased gene; Unb, unbiased gene. Numbers in brackets represent the number of genes. The lower and upper hinges of the box correspond to the first and third quartiles (the 25th and 75th percentiles). The upper whisker extends from the hinge to the largest and smallest values no further than 1.5x IQR from the hinge (Inter-Quartile Range, distance between the first and third quartiles). *P*-values represent a two-sample Wilcoxon rank sum test.

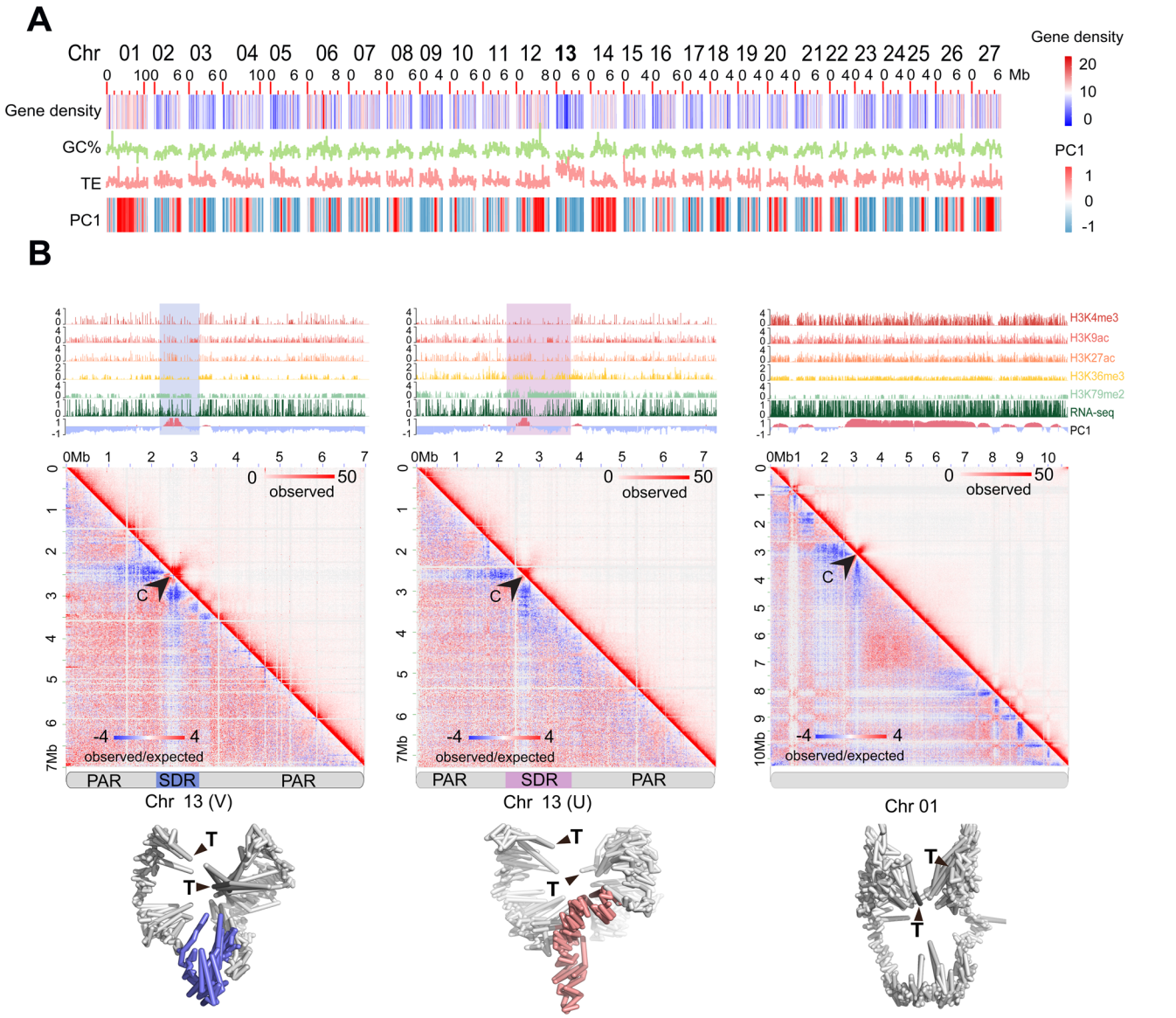



**Fig. 4 | U and V sex chromosome 3D architecture. A** Plot showing gene density, GC content, and TE density in 100 kb windows and compartment A/B (PC1, red indicates positive values corresponding to compartment A, and blue indicates negative values corresponding to compartment B) in 10 kb windows across the *Ectocarpus* chromosomes. Chromosome 13 is the sex chromosome. **B** Hi-C map and simulated 3D configurations of sex chromosomes at 10 k resolution, employing a maximum likelihood approach, chromosomal structures were constructed from Hi-C data with a default setting of 3D Max[105]. SDRs in the simulated male and female chromosomes are colored in blue and red, respectively, and telomeres are labeled with black triangles. In each panel, the black arrowhead indicates the centromere. The tracks above each Hi-C map show A/B compartment annotation (PC1), gene expression (RNA-seq), and various histone modification ChIP-seq.

https://doi.org/10.1038/s41467-024-53453-5

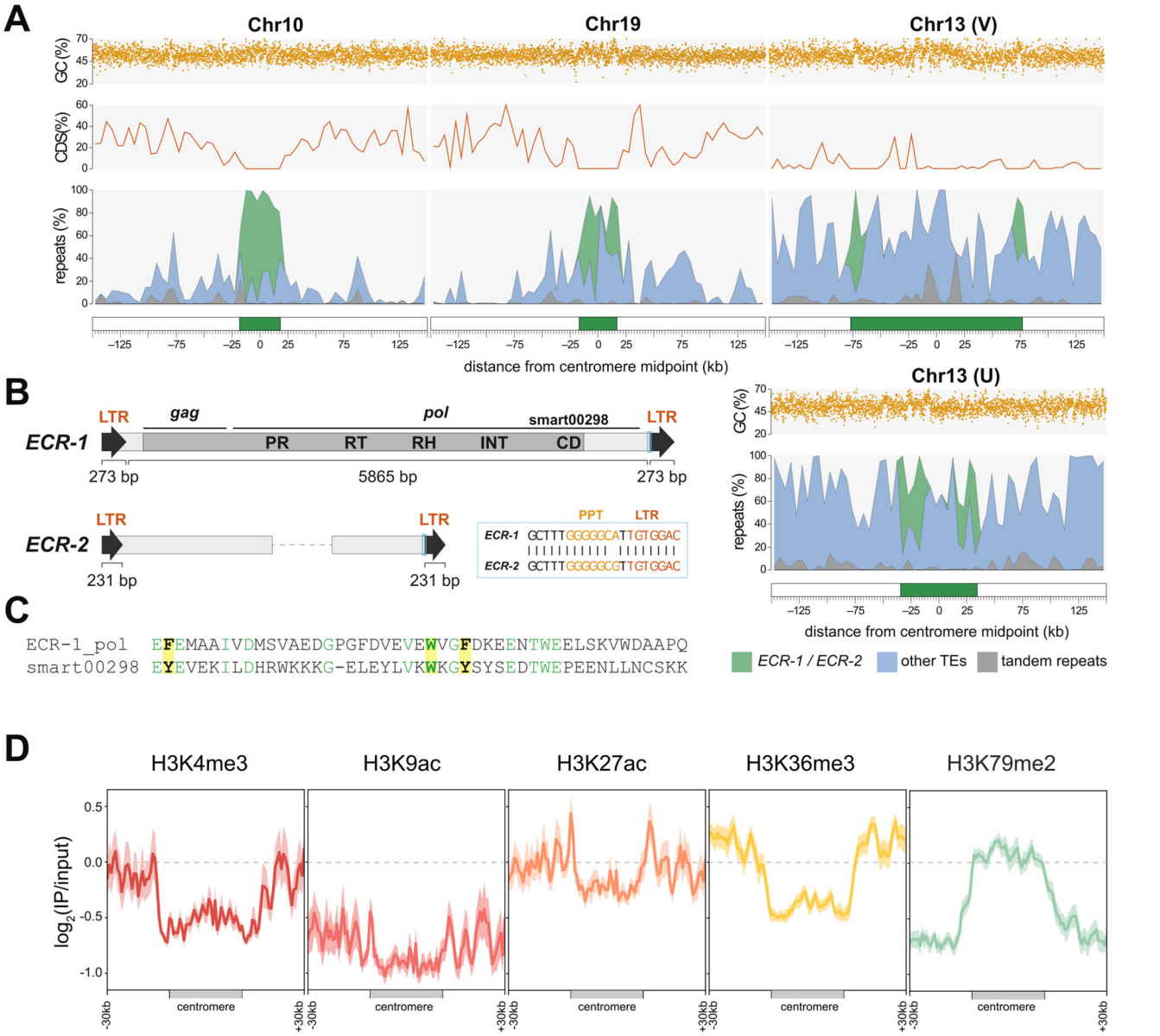



**Fig. 5 | *Ectocarpus* centromeres and centromere-specific retrotransposons.** **A** The centromeric regions of select chromosomes and the *ECR* retrotransposons. Putative centromeres and flanking regions for four chromosomes, including the U chromosome from the female genome assembly. The centromere (green box) is defined as the region from the first to the last copy of *ECR* elements. The repeats panel is shown as a stacked area plot, and the percentage of each repeat type is plotted in 5 kb windows. Coding sequence (CDS) density is plotted in 5 kb windows, and GC content is plotted in 100 bp windows. For all chromosomes, see Supplementary Fig. 9, and for genomic coordinates, see Supplementary Data 6. **B** Schematics of the *ECR* retrotransposons. The light blue boxes highlight the conserved region between *ECR-1* and *ECR-2*, and a partial alignment of this region is shown (PPT = polypurine tract). Only 5′ and 3′ fragments of *ECR-2* were recovered, and the dashed line represents a protein-coding sequence that is presumably missing. The domains shown on the *ECR-1* protein are: PR = protease, RT = reverse transcriptase, RH = RNaseH, INT = integrase, CD = chromodomain. **C** Alignment of *ECR-1* chromodomain and SMART chromodomain curated model (smart00298). Conserved amino acids are colored green, and the three aromatic amino acids that are responsible for the recognition of histone-methylated lysines are highlighted in yellow. **D** Histone mark signal (log2(IP/input)) in the putative centromeres and the surrounding regions (30 kb). Profiles of histone marks around the centromeres. The solid line represents the $\log_2$ of the ChIP-seq signal relative to the input, while the shading represents the 95% confidence interval (CI) of the center for the error bands. Heatmaps showing the chromatin state of each centromere using both uniquely and multi-mapped reads are shown in Supplementary Fig. 10.

share a ~ 64 bp region of homology that includes the polypurine tract that immediately precedes the 3′ terminal repeat (Fig. 5B). We name these elements *ECR-1* and *ECR-2* for *Ectocarpus* Centromeric Retrotransposon.

Notably, the *ECR-1* polyprotein features a C-terminal chromodomain fused to the integrase domain (Fig. 5B, C). Chromodomains recognize and bind histone-methylated lysines via a cage tertiary structure that is formed by three aromatic residues[52], all of which are conserved in the *ECR-1* polyprotein (Fig. 5C).

Defining the putative centromeres as the region between the first and last *ECR* element, lengths range from only 6.8 kb on chromosome 25 (essentially a single copy of *ECR-1*) to 153.9 kb on the male chromosome 13 (i.e., chromosome V), with a median of 38.6 kb (Supplementary Data 6). On average per centromere, 33% of bases are contributed by *ECR-1*, 7.3% by *ECR-2*, and 34% by other interspersed repeats that are not exclusive to these regions. Tandem repeats constitute only 3.5% of the putative centromeres, relative to 6.7% elsewhere in the genome. Although genes were generally absent from

these regions, certain chromosomes feature a small number of genes distributed among the *ECR* copies, including chromosome V, (Fig. 5A). Furthermore, only one of the four SDR genes that are contained within the V centromere was sex-linked in the last brown algal common ancestor (Ec-13_001830), while the others became sex-linked later during the expansion of the SDR in the Ectocarpales (Ec-13_001830.1) and in the expansion that is exclusive of *Ectocarpus* (Ec-13_001870.1 and Ec-13_001890.1)[49]. The GC content of the putative centromeres (52.7%) is only marginally lower than the rest of the genome (53.5%). However, this is partly driven by the GC content of *ECR-1* (58.2%), and several chromosomes do feature short AT-rich sequences within the putative centromeres (e.g., chromosome 19, Fig. 5A). As expected, following their evolutionary independence, the putative centromere of the female U chromosome differs substantially in length and composition relative to the V chromosome.

To further characterize the *Ectocarpus* putative centromeres, we analyzed the associated chromatin pattern using ChIP-seq data (Fig. 5D and Supplementary Fig. 10). Despite their assignment to compartment A, the putative centromeres exhibit a slight enrichment with the H3K79me2 mark in comparison to the surrounding genomic regions in the male SDR. In contrast, the histone marks associated with active genes (H3K4me3, H3K9ac, H3K27ac, and H3K36me3) were strongly depleted within the putative centromeres but strongly enriched in the surrounding regions, consistent with their compartment assignment and the presence of flanking genes. Interestingly, on a few chromosomes, the H3K79me2 pattern extends beyond the boundaries of the *ECR* elements. This observation holds true when using a different mapping method (removing multi-mapping reads) for chromosomes 16, 22, and V (Supplementary Fig. 10A, B), all of which have flanking regions that are highly enriched with interspersed repeats (TEs).

### An inserted (endogenous) viral element exhibits a unique chromatin conformation

Marine filamentous brown algae of the order Ectocarpales frequently carry endogenous giant viruses with large double-stranded DNA genomes[53]. *Ectocarpus sp. 7*, in particular, has been shown to harbor such type of endogenous viral element inserted in chromosome 6, derived from the *Ectocarpus* phaeovirus EsV-1[31,54,55]. We confirmed the presence of an endogenous viral element (that we name Ec32EVE) localized within chromosome 6 in our V5 *Ectocarpus* genome (Fig. 6A, see also Fig. 1A,). The Ec32EVE is 399 kbp long, contains 199 genes, and is covered with a large domain of the repression-associated mark H3K79me2 previously shown to be associated with the silencing of transposable elements in *Ectocarpus*[42,43]. The Ec32EVE region exhibits a depletion of activation-associated histone marks H3K4me3, H3K9ac, H3K27ac, and H3K36me3 (Fig. 6A). Consistent with this heterochromatic landscape, RNAseq analysis showed negligible expression throughout the entire Ec32EVE region (Fig. 6A), highlighting the silent nature of the potentially coding regions within the endogenous viral element. The chromosome 6 Hi-C map further revealed high levels of compaction and insulation, associated with the viral insertion region (Fig. 6A). Remarkably, the Ec32EVE region displayed strong long-range contact with telomeres in the nuclear 3D space (Fig. 6B). We asked whether this observation was related to the highly heterochromatic nature of the Ec32EVE region. However, other genomic regions equally marked with long stretches of H3K79me2 did not necessarily cluster in 3D with telomeres (Supplementary Fig. 11). It appears, therefore, that the Ec32EVE insertion, rather than the chromatin state of this region per se, is implicated in the unique 3D structure of this region.

## Discussion

High-quality and complete reference genome assemblies are fundamental for the application of genomics to a range of disciplines in biology, from evolutionary genomics genetics to biodiversity conservation. Here, we obtained a highly accurate and nearly complete assembly of the reference genome of the brown alga *Ectocarpus*, a model organism for this key group of eukaryotes. The *Ectocarpus* V5 assembly includes telomeres for most chromosomes and very few gaps and, therefore, provides a new reference genome for the scientific community.

Chromosome folding patterns vary across lineages[56]. For example, in many plant species with relatively large genomes, chromosomes adopt a Rabl configuration during interphase, in which centromere or telomere bundles are associated with opposite faces of the nuclear envelope. For chromosomes with Rabl configuration, their Hi-C maps display a characteristic belt that is perpendicular to the primary diagonal. *Arabidopsis*, in contrast, presents a Rosette configuration[57], where the Hi-C maps feature conspicuous long-range intra-chromosomal contacts due to the formation of megabase-size loops. None of these features were found in the *Ectocarpus* Hi-C map, suggesting that its chromatin adopts a non-Rabl and non-Rosette configuration. The Chromatin arrangement of interphase chromosomes in *Ectocarpus* involved telomeres of all chromosomes and centromeres of all chromosomes clustering together. Therefore, despite different linear genome architectures and centromere sequence compositions, centromere interactions appear to be a pervasive feature in eukaryotes, from plants and animals to brown algae.

TADs, whose boundaries partition the genome into distinct regulatory territories, are a prevalent structural feature of genome packing in animal and plant species, but our observations showed that TADs are not prominent in the *Ectocarpus* genome. Note that *Ectocarpus* has a relatively simple morphology with a reduced number of cell types. The Hi-C maps, thus, are likely to faithfully represent the interphase chromatin structure of male and female *Ectocarpus* rather than an average conformation across multi-cell types as in other more complex organisms. This feature allows us to conclude that *Ectocarpus* has a non-Rabl chromatin conformation and does not exhibit TADs at a local level. The *Arabidopsis* genome is another example in which TADs are absent[26], and this feature is thought to be related to *Arabidopsis* small genome size, high gene density, and short intergenic regions. Given that *Ectocarpus* has similar genomic characteristics, the absence of TADs in *Ectocarpus* supports the hypothesis that TADs may form when the genome size is above a certain threshold[58]. Note that in *Arabidopsis*, despite its genome not having clear TADs, over 1000 TAD-boundary-like and insulator-like sequences were found from Hi-C maps normalized with 2 kb genomic bins[26]. These regions possess similar properties to those of animal TAD borders/insulators, i.e., chromatin contacts crossing insulator-like regions are restricted, and they are enriched for open chromatin. The *Ectocarpus* genome, in contrast, is mainly partitioned in H3K79me2 rich and H3K79me2 poor regions, that largely define A and B compartments, but we did not find any evidence for canonical insulators nor 'TAD-boundery-like' regions. Note that CTCF is absent in the genome of *Ectocarpus*, similar to yeast, *Caenorhabditis elegant,* and plants[59].

The high-resolution, sex-specific Hi-C maps of haploid individuals allowed us to examine the 3D structure of the U and V sex chromosomes in the interphase nucleus of *Ectocarpus* males and females. The U and V chromosomes are largely homomorphic, each containing a small, non-recombining region[9,49] that harbors several dozen genes, including the master male-determining factor *MIN*[60], and a largely heterochromatic landscape[43]. Our *Ectocarpus* V5 yielded gapless SDRs and demonstrated that the U and V SDRs span the centromere. The linkage between the mating type (MT) locus and centromeres is a common feature of haploid MT chromosomes in fungi. For example, in the *Microbotryum* fungi, recombination suppression links the MT-determining loci to centromeres[61,62], and this is thought to help preserve heterozygosity and/or be beneficial under auto-fecundation, as it increases the degree of compatibility between gametes from the same individual. A similar process is unlikely to be operating in *Ectocarpus* because there is no intra-tetrad direct crossing; haploid spores

https://doi.org/10.1038/s41467-024-53453-5

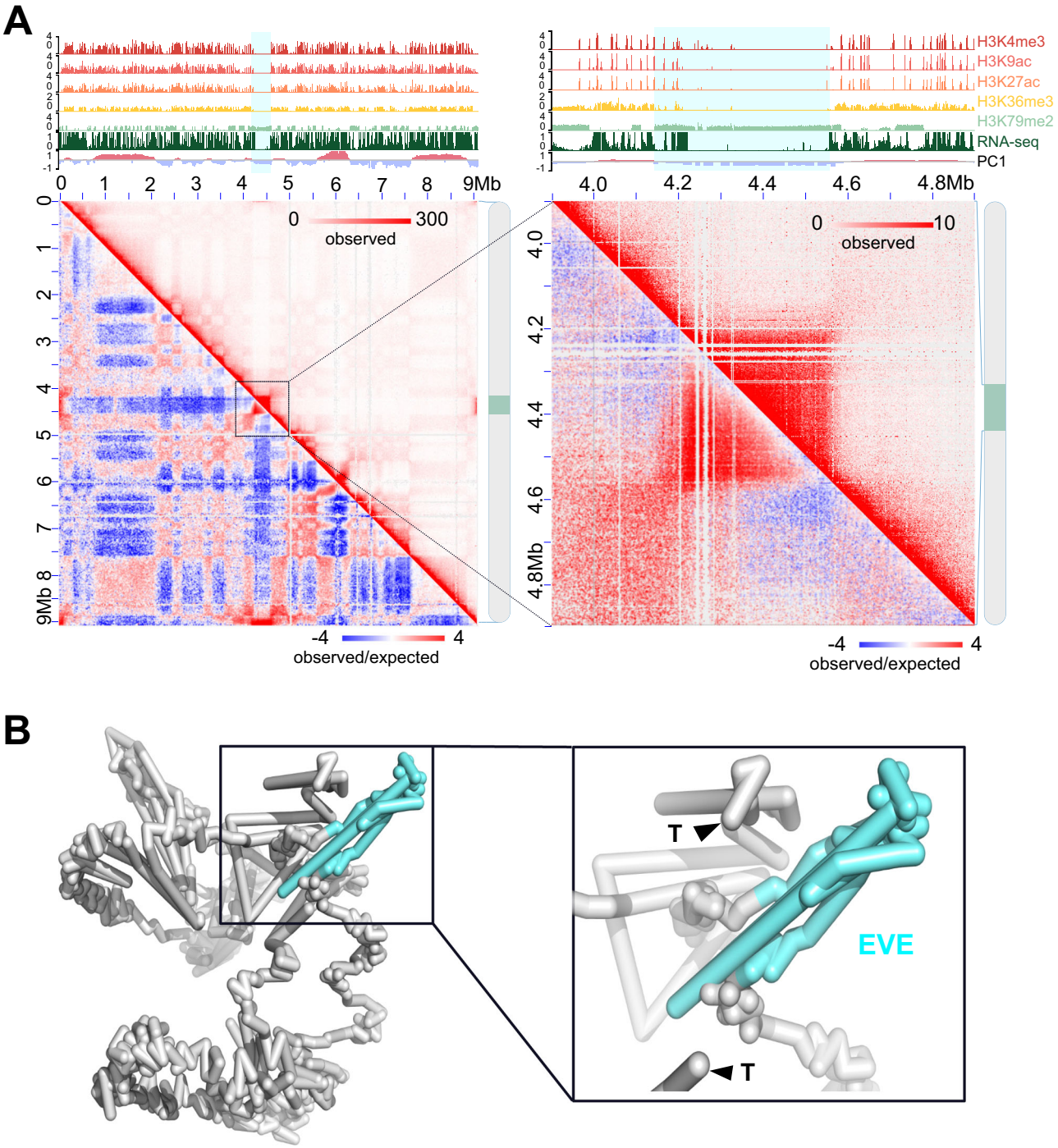



**Fig. 6 | Virus insertion region (EVE) is insulated and shows strong interactions with telomeres. A** Hi-C map of chromosome 6. The zoomed-in region to the right contains the EVE (4.2–4.6 Mb). The tracks above each Hi-C map show A/B compartment annotation (PC1), gene expression (RNA-seq), and histone PTMs ChIP-seq tracks. A blue shade marks the region of the Ec32EVE. **B** Simulated 3D configurations of chromosome 6 at 10 k resolution, generated using 3DMax[105]. The EVE region is colored in aquamarine, and the telomeres are labeled with black triangles. To the right, the Ec32EVE region is zoomed in to highlight the long-range contacts with the telomeres.

disperse after meiosis, develop into male and female gametophytes, and produce gametes at a later stage[63]. Furthermore, only one of the four SDR genes that are contained within the V centromere was sex-linked in the last brown algal common ancestor (Ec-13_001830), while the others became sex-linked later during the expansion of the SDR in the Ectocarpales (Ec-13_001830.1) and in the expansion that is exclusive of *Ectocarpus* (Ec-13_001870.1 and Ec-13_001890.1)[49]. It is, therefore, more conceivable that SDR linkage to the centromere in *Ectocarpus* occurred due to expansion of the non-recombining SDR, during which the centromere was subsumed in this region likely via a large-scale inversion, as suggested by comparative genomic studies[49]. Nonetheless, we cannot fully exclude a scenario where the ancient SDR genes were initially located in the centromere and were later translocated elsewhere as the SDR expanded.

What is the potential role of the sex chromosome 3D chromatin configuration? Among numerous steps required for gene expression, the spatial organization of the genome is known to modulate DNA accessibility to the transcriptional machinery and to promote contacts

between genes and distant regulatory DNA elements such as enhancers[25]. In the case of *Ectocarpus*, the correct spatial and temporal window of transcriptional activation of genes contained within the SDR is critical to ensure sex determination and differentiation in the brown algal tissues. It is, therefore, likely that the tight transcriptional regulation of the SDR is achieved both by 3D chromatin remodeling in conjunction with histone PTMs and small RNAs[44]. Whilst the 3D chromatin configuration of animal and plant sex chromosomes remains largely elusive, it is well known that chromatin 3D structure is involved in the repression of the silent MT loci in yeast *Saccharomyces cerevisiae* during mating type switching[64,65]. Therefore, it appears that modulation of mating type or sex chromosome architecture may play a significant role in controlling sex-specific features across eukaryotic lineages.

The *Ectocarpus* V5 assembly and high-resolution Hi-C map allowed us to examine centromeric sequences in this organism. We observed 27 unique centromere sequences occurring once per chromosome, a finding that helps to resolve nuclear genome organization and indicates monocentric regional centromeres. The centromeres of *Ectocarpus* may be categorized as transposon-rich and primarily composed of centromere-specific retrotransposons, a relatively common centromeric organization found in species including the amoeba *Dictyostelium discoideum*[66,67] and the oomycete *Phytophthora sojae*[68], where the centromere-specific retrotransposons are associated with the centromere-specific histone H3 variant (cenH3). Although *ECR-1* presumably targets centromeric DNA (and *ECR-2* may have done so in the past), it remains to be determined whether the *ECR* elements constitute the centromere in *Ectocarpus*. The putative centromeres are short relative to the transposon-rich centromeres of many other species[51], and we cannot rule out an association between cenH3 and short AT-rich sequences, as in diatoms[69]. ChIP-sequencing of cenH3 will be required to distinguish between these possibilities.

The fusion of an LTR integrase to a C-terminal chromodomain is most widely known from the evolutionary ancient chromovirus clade of *Metaviridae* LTRs, where the presence of the chromodomain enables recognition of specific histone modifications and targeted insertion at associated genomic sites[70]. In plants, the CRM subclade of chromovirus LTRs contains many centromere-targeting families that accompany satellite arrays and constitute a major component of centromeric DNA[71]. Independent lineages of chromodomain-containing LTRs have been reported in Stramenopiles, including the Chronos *Metaviridae* elements of oomycetes[72] and the *CoDi*-like *Pseudoviridae* (i.e., *Ty1*/*Copia*) elements of diatoms[73]. Interestingly, *ECR-1* does not appear to be a member of either the chromovirus or Chronos clades and instead is most closely related to chromodomain-containing oomycete LTRs that are yet to be phylogenetically classified (e.g., *Gypsy-20_PR* from *Phytophthora ramorum*). We hypothesize that the chromodomain of *ECR-1* may enable centromere-targeted integration in *Ectocarpus*, either by recognition of cenH3 or other centromere-associated proteins, implying evolutionary convergence with the CRM elements of plants. However, it is unlikely that the targeting mechanism is itself convergent since the centromere-targeting CRM elements feature derived chromodomains that lack the three conserved aromatic amino acids[71], which are present in *ECR-1*.

Viruses that transcribe their DNA within the nucleus have to adapt to the molecular mechanisms that govern transcriptional regulation. The interaction between chromatin and viral-directed modulation of chromatin is a critical component of the viral-host interaction[74]. However, the complexity of the higher-order organization of the host genome and its potential influence in the regulation of gene expression raises questions regarding the spatial arrangement of integrated viral DNA in the host's genome. Phaeoviruses are latent giant double-stranded DNA viruses that insert their genomes into those of their brown algal (Phaeophyceae) hosts[53,55]. Remarkably, although about 50% of individuals in *Ectocarpus* field populations show symptoms of giant viral infection[75], the *Ectocarpus* strain used in this study has never been observed to produce virus particles, and Ec32EVE genes are transcriptionally silent[31]. Here, we showed that the silencing of Ec32EVE genes correlated with the deposition of large domains of repressive-associated chromatin mark H3K79me2, concomitant with depletion of activation-associated marks. Moreover, the inserted giant viral element was associated with the B compartment, and adopted a highly insulated conformation in the 3D nuclear space, exhibiting strong long-range contacts with the telomeres. It is possible that mechanisms such as phase separation and maybe loop extrusion[76] may underlie the specific 3D configuration of this region. Whilst the detailed mechanisms underlying the relationship between giant virus latency and gene-silencing mechanisms, including the 3D architecture of the chromatin, remain to be determined, our study provides the first description of the 3D configuration of an inserted giant viral element and strong evidence for an interplay between 3D chromatin architecture, H3K79me2 domains, and EVE gene silencing, opening new avenues to gain insights regarding the functional significance of these interactions.

## Methods

### Brown algae culture

Algae were cultured as previously described[77]. Briefly, *Ectocarpus* strains Ec32, Ec25, Ec561, and Ec560 were grown in autoclaved natural seawater (NSW) with PES at 14 °C with the light intensity of 20 µmol photons $m^{-2}$ $s^{-1}$ (12 h light/12 h dark). The medium was changed every week. Before collection, algae were treated with antibiotics: Streptomycin (25 mg/L), Chloramphenicol (5 mg/L), and PenicillinG (100 mg/L) for three days to limit bacterial growth.

### Hi-C

An in situ Hi-C protocol of plants[78] was optimized for brown algae. *Ectocarpus* cultures were collected using a 40 µm filter and fixed in 2% (vol/vol) formaldehyde for 30 min at room temperature, and the cross-linking reaction was quenched with 400 mM glycine. Approximately 50 mg fixed algae suspended in 1 ml nuclei isolation buffer (0.1% triton X-100, 125 mM sorbitol, 20 mM potassium citrate, 30 mM MgCl2, 5 mM EDTA, 5 mM 2-mercaptoethanol, 55 mM HEPES at pH 7.5) with 1X protease-inhibitor in a 2 ml VK05 tube, then homogenized by Precellys Evolution beads homogenizer (Bertin technologies) with the following settings: 7800 rpm, 30 s each time, 20 s pause each grinding cycle, repeat 5 times. Over 1 million nuclei were isolated and digested overnight by Dpn II, DNA ends were labeled with biotin-14-dCTP, then ligated by T4 DNA ligase enzyme. The purified Hi-C DNA was sheared by covaries E220 evolution and libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit (NEB, no. E7645), and the average size of the library was detected by bioanalyzer, the final library was sequenced with 150 bp paired-end reads on an Illumina HiSeq 3000 platform. Two biological replicates were performed for each strain.

### Nanopore sequencing

High molecule weight (HMW) DNA of *Ectocarpus* male (Ec32) and female (Ec25) were isolated using *OmniPrep*™ kit (G-Biosciences) with slight modifications. 500 mg of fresh collected tissue was dried and resuspended in 1 ml lysis buffer, then homogenized using a Precellys mixer. Samples were incubated at 60 °C for 1 h with proteinase K, inversed every 15 min. HMW-gDNA was dried and eluted by 10 mM ph 8.0 Tris-HCl, and incubated at 55 °C for 30 min with 0.5 µL 10 mg/ml RNaseA. The concentration of HMW-gDNA was quantified using an Invitrogen Qubit 4 Fluorometer, and molecule size distributions were estimated using a FEMTO Pulse system (Agilent). The sample was further cleaned and concentrated using AMPure XP SPRI paramagnetic beads (Beckman Colter) at a DNA: bead volume ratio of 1:0.6, followed by two washes using freshly prepared 70% ethanol and resuspension in 10 mM ph 8.0 Tris-HCl. 1 µg HMW-gDNA was used for nanopore library

preparation and sequencing according to the standard protocol of the ONT Ligation Sequencing Kit (Nanopore, https://store.nanoporetech.com/eu/ligation-sequencing-kit110.html). Sequencing was performed on an ONT MinION Mk1B with three R9.4.1 flow cells.

### Re-assembly of genomes assisted by Nanopore and Hi-C

Base-calling was done by ONT Guppy v6.5.7 (--trim_adapters –trim_primers)(Wick et al., 2019). A de novo draft male genome assembly was generated based on Ec32 ONT data by the Canu assembler v2.2(genome Size = 220 m -pacbio-raw)(Koren et al., 2017), with three iterations of error correction by Pilon v1.24[79]. An additional scaffolding step was accomplished by ARCS v1.2.5 ($z = 1500$ $m = 8$-$10000$ $s = 70$ $c = 3$ $l = 3$ $a = 0.3$)[80]. As long read sequencing input, the original ONT read data was extended by the previous assembly[31]; the same strategy was used for the new *Ectocarpus* female draft chromosome 13 with Ec25 ONT reads only mapped to the male chromosome 13 and previous published female SDR scaffold.

The Hi-C raw reads underwent a preprocessing step using Trimmomatic v.0.39 with a default setting to remove the adapters and other Illumina-specific sequences[81]. Subsequently, the clean reads were aligned draft genomes using a 3D de novo assembly (3D-DNA) pipeline, following[82]. The resulting Hi-C contact map, based on the initial chromosomal assembly, was visualized using Juicebox[83]. Juicebox also facilitated the manual adjustment of contig orientations and order along the chromosomes, based on the observed contacts. During this adjustment process, some incorrectly placed sequences were trimmed from the original contigs and reassembled with the appropriate ones. The orientation of the final chromosome name was corrected with the previous reference genome[30]. To refine the assembly, we employed TGS-GapCloser with error correction by racon v1.4.3, along with RFfiller utilizing ONT reads for gap filling[84,85]. Subsequently, an assessment of genome quality was conducted by Benchmarking Universal Single-Copy Orthologs (BUSCO)[34] together with its eukaryote and stramenopiles databases in version odb10.

To be consistent with the V2 genome, we extracted the gapless 1.55 Mb female sex-determining region (SDR) of the female assembly and added it as a separate contig to the male genome (fSDR). This 'reference' assembly is the new *Ectocarpus sp* V5 reference genome.

To identify bacterial contamination in the genome assemblies, the newly assembled scaffolds were analyzed by kraken2 (version 2.1.3)[86], blastn (version 2.13.0, nt database 2022-07-01)[87] and blob tools (version 1.1.1)[88]. Hits identified by all three tools were considered, and corresponding contamination scaffolds were removed. During the contamination analyses, we removed two Hi-C scaffolds corresponding to the bacteria genera *Paraglaciecola* and *Halomonas*.

### Hi-C data analysis

The Hi-C reads were processed using the Juicer pipeline[89], and binning was performed at various sizes, including 2, 5, 10, 20, 50, 100, and 500 kb. The clean Hi-C data was mapped to its corresponding reassembled reference genome (male or female *Ectocarpus* V5, Supplementary Fig. 1) using Bowtie2[90]. During the alignment, the clean reads were aligned end-to-end, and spanning ligation junctions were trimmed at their 3'-end and realigned to *Ectocarpus* newly assembled genome. The resulting aligned reads from both fragment mates were then paired and stored in a paired-end BAM file. Invalid Hi-C reads, including discarding dangling-end reads, same-fragment reads, self-circled reads, and self-ligation reads, were removed from further analyses.

### Chromosomal contact probability

The reads information processed by the Juicer pipeline in the "merged_nodups.txt" were converted to pairs using the pairix tool[91]. The draft genome was divided into 1000 bp bins, and the contact probability $P(s)$ was calculated and visualized using cooltools[92] following the guidelines provided in the documentation at https://cooltools.readthedocs.io/en/latest/notebooks/contacts_vs_distance.html. In short, $P(s)$ was determined by dividing the number of observed interactions within each bin by the total number of possible pairs.

### A/B compartment identification

The A/B compartment status was determined using Eigenvalues (E1) obtained through eigenvector decomposition of Hi-C contact maps. To calculate the E1 values at a 10 kb resolution, Cooltools software was utilized with the "cooltools eigs-trans" function and GC density file[92]. The resulting E1 values were then loaded into the plaid pattern of Hi-C contact maps. Manual validation based on intra or inter-chromosomal interactions in Hi-C was performed along each chromosome to obtain the final list of "E1" values. Since the direction of eigenvalues is arbitrary, positive values were assigned the label "A", while negative values were assigned the label "B" based on their association with GC or gene density. The compartment border was defined as the edge bin separating the A and B compartments.

### ChIP-seq and RNA-seq

ChIP-seq and RNA-seq data from the male (Ec561) and female (Ec560) strains were obtained from[44]. The datasets include two replicates of H3K4me3, H3K9ac, H3K27ac, H3K36me3, and H3K79me2 samples, as well as two control samples (an input control corresponding to sonicated DNA and anti-histone H3). To process the data, the nf-core ChIP-seq pipeline v2.0.0 was employed[93]. Briefly, the raw data underwent trimming using Trim Galore v0.6.4[94], and the paired-end reads were aligned to the reference genome using BWA v0.7.17[95]. Subsequently, MACS2 with default parameters was used to call broad and narrow peaks[96]. Peaks called with MACS2 for each of the histone marks (normalized by H3) were compared in the A versus B compartment regions to examine the enrichment of each mark per compartment. In short, we used MACS2 to call peaks for each IP, with H3 as the control. The 'fold_enrichment' values from the generated files were used for co-analysis with compartments A/B. The plot values (represented on the *y*-axis) are the log2-transformed 'fold_enrichment' values located in each compartment A or B.

Three replicates of RNA-seq data were trimmed by Trimmomatic v0.39 and mapped on the *Ectocarpus* V5 reference genome (Supplementary Fig. 1) by GSNAP aligner v2021-12-17[81,97], unique mapped read pairs were used to calculate read counts per gene by featureCount v2.0.3, DEseq2 (v1.41.6, Bioconductor) was used for detection differential expression genes with the threshold of adjusted $p$-value = < 0.01 and log2fold change > = 1, TPM (Transcripts Per Million) was used for transcript abundance quantification[43,98,99].

### Centromere characterization

Broad centromeric regions were determined by visually assessing the Hi-C contact maps. To assess the repeat content of these regions, RepeatModeler v2.0.2[100] was run on the male Ec32 V5 genome assembly to generate de novo repeat consensus models, using the flag "-LTRStruct" to perform LTR structural searches. The subsequent repeat library was provided as input to RepeatMasker v4.0.9 (https://www.repeatmasker.org/RepeatMasker/) to identify the genomic coordinates of repeats. Tandem Repeats Finder v4.09.1[101] was run to identify coordinates of satellite and microsatellite DNA using the recommended parameters "2 5 7 80 10 50 2000", enabling satellite DNA with monomers up to 2 kb to be identified. Final tandem repeat coordinates were achieved by combining the simple and low-complexity repeats identified by RepeatMasker with the repeats identified by Tandem Repeats Finder. All other repetitive coordinates identified by RepeatMasker that did not overlap tandem repeats were assumed to be interspersed repeats (i.e., transposable elements).

Putative centromeric repeats were identified by searching for repeat families that were both almost exclusively present in the broad

centromeric regions defined by the contact maps and common to all chromosomes. The two repeat models that met these criteria were then manually curated following Goubert et al.[102]. Retrotransposons related to *ECR-1* were identified by passing the predicted protein to Repbase Censor online tool[103]. Centromeric coordinates were defined as the first to the last copy of *ECR* elements (see Supplementary Data 6). All centromeric analyses were performed on the male Ec32 V5 genome, except for the U chromosome, which was analyzed using the female Ec25 genome.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

The Nanopore and Hi-C data generated in this study have been deposited in the NCBI database under the project number PRJNA1105946. *Ectocarpus* V5 genome and gene annotation, the processed Hi-C, ChIP-seq, and RNA-seq data are available in Edmond of Max Planck Digital Library collection (https://doi.org/10.17617/3.QXUAMN and https://doi.org/10.17617/3.NQDSLW). The RNA-seq and ChIP-seq datasets used in this study were retrieved from the NCBI Gene Expression Omnibus repository: PRJNA1055718.

## References


1. Bachtrog, D. et al. Are all sex chromosomes created equal? *Trends Genet.* **27**, 350–357 (2011).
2. Charlesworth, D. Plant Sex Chromosomes. *Annu. Rev. Plant Biol.* **67**, 397–420 (2016).
3. Abbott, J. K., Nordén, A. K. & Hansson, B. Sex chromosome evolution: historical insights and future perspectives. *Proc. R. Soc. B Biol. Sci.* **284**, 20162806 (2017).
4. Ponnikas, S., Sigeman, H., Abbott, J. K. & Hansson, B. Why do sex chromosomes stop recombining? *Trends Genet.* **34**, 492–503 (2018).
5. Umen, J. & Coelho, S. Algal sex determination and the evolution of anisogamy. *Annu. Rev. Microbiol.* **73**, 267–291 (2019).
6. Olito, C. & Abbott, J. K. The evolution of suppressed recombination between sex chromosomes and the lengths of evolutionary strata. *Evolution* **77**, 1077–1090 (2023).
7. Charlesworth, B. Model for evolution of Y chromosomes and dosage compensation. *Proc. Natl. Acad. Sci. USA* **75**, 5618–5622 (1978).
8. Beukeboom, L. & Perrin, N. *The Evolution of Sex Determination*. (Oxford University Press, 2015).
9. Ahmed, S. et al. A haploid system of sex determination in the brown alga *Ectocarpus* sp. *Curr. Biol. CB* **24**, 1945–1957 (2014).
10. Mahajan, S., Wei, K. H.-C., Nalley, M. J., Gibilisco, L. & Bachtrog, D. De novo assembly of a young Drosophila Y chromosome using single-molecule sequencing and chromatin conformation capture. *PLOS Biol.* **16**, e2006348 (2018).
11. Cechova, M. et al. Dynamic evolution of great ape Y chromosomes. *Proc. Natl. Acad. Sci. USA* **117**, 26273–26280 (2020).
12. Montgomery, S. A. et al. Chromatin organization in early land plants reveals an ancestral association between H3K27me3, transposons, and constitutive heterochromatin. *Curr. Biol. CB* **30**, 573–588 (2020).
13. Carey, S. B. et al. Gene-rich UV sex chromosomes harbor conserved regulators of sexual development. *Sci. Adv.* **7**, https://doi.org/10.1126/sciadv.abh2488 (2021).
14. Moraga, C. et al. The *Silene latifolia* genome and its giant Y chromosome. Preprint at https://doi.org/10.1101/2023.09.21.558754 (2023).
15. Rhie, A. et al. The complete sequence of a human Y chromosome. *Nature* **621**, 344–354 (2023).
16. Yue, J. et al. The origin and evolution of sex chromosomes, revealed by sequencing of the Silene latifolia female genome. *Curr. Biol.* **33**, 2504–2514 (2023).
17. Misteli, T. The self-organizing genome: Principles of genome architecture and function. *Cell* **183**, 28–45 (2020).
18. Chen, L.-F. & Long, H. K. Topology regulatory elements: From shaping genome architecture to gene regulation. *Curr. Opin. Struct. Biol.* **83**, 102723 (2023).
19. Rowley, M. J. et al. Evolutionarily conserved principles predict 3D chromatin organization. *Mol. Cell* **67**, 837–852 (2017).
20. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
21. Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci.* USA **112**, https://doi.org/10.1073/pnas.1518552112 (2015).
22. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 (2017).
23. Dong, P. et al. 3D Chromatin architecture of large plant genomes determined by local A/B compartments. *Mol. Plant* **10**, 1497–1509 (2017).
24. Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E. & Wiehe, T. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc. Natl. Acad. Sci. USA* **109**, 17507–17512 (2012).
25. Szabo, Q. et al. Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nat. Genet.* **52**, 1151–1157 (2020).
26. Wang, C. et al. Genome-wide analysis of local chromatin packing in *Arabidopsis thaliana*. *Genome Res.* **25**, 246–256 (2015).
27. Crevillén, P., Sonmez, C., Wu, Z. & Dean, C. A gene loop containing the floral repressor FLC is disrupted in the early phase of vernalization. *EMBO J.* **32**, 140–148 (2012).
28. Ramirez-Prado, J. S., Rodriguez-Granados, N. Y., Ariel, F., Raynaud, C. & Benhamed, M. Chromatin architecture: A new dimension in the dynamic control of gene expression. *Plant Signal. Behav.* **11**, e1232224 (2016).
29. Lee, H. & Seo, P. J. Accessible gene borders establish a core structural unit for chromatin architecture in Arabidopsis. *Nucleic Acids Res.* **51**, 10261–10277 (2023).
30. Cormier, A. et al. Re-annotation, improved large-scale assembly and establishment of a catalogue of noncoding loci for the genome of the model brown alga *Ectocarpus*. *N. Phytol.* **214**, 219–232 (2017).
31. Cock, J. M. et al. The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* **465**, 617–621 (2010).
32. Chaux-Jukic, F. et al. Architecture and evolution of subtelomeres in the unicellular green alga *Chlamydomonas reinhardtii*. *Nucleic Acids Res.* **49**, 7571–7587 (2021).
33. Kawai, H., Nakayama, T., Inouye, I. & Kato, A. LINKAGE OF 5S RIBOSOMAL DNA TO OTHER rDNAS IN THE CHROMOPHYTIC ALGAE AND RELATED TAXA [1]. *J. Phycol.* **33**, 505–511 (1997).
34. Manni, M., Berkeley, M. R., Seppey, M. & Zdobnov, E. M. BUSCO: Assessing genomic data quality and beyond. *Curr. Protoc.* **1**, https://doi.org/10.1002/cpz1.323 (2021).
35. Ouyang, W., Xiong, D., Li, G. & Li, X. Unraveling the 3D genome architecture in plants: Present and future. *Mol. Plant* **13**, 1676–1693 (2020).
36. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
37. Grob, S., Schmid, M. W. & Grossniklaus, U. Hi-C Analysis in arabidopsis identifies the KNOT, a structure with similarities

to the flamenco locus of Drosophila. *Mol. Cell* **55**, 678–693 (2014).
38. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
39. Ruiz-Velasco, M. & Zaugg, J. B. Structure meets function: How chromatin organisation conveys functionality. *Curr. Opin. Syst. Biol.* **1**, 129–136 (2017).
40. Yin, X. et al. Binding by the Polycomb complex component BMI1 and H2A monoubiquitination shape local and long-range interactions in the Arabidopsis genome. *Plant Cell* **35**, 2484–2503 (2023).
41. Pei, L., Li, G., Lindsey, K., Zhang, X. & Wang, M. Plant 3D genomics: the exploration and application of chromatin organization. *N. Phytol.* **230**, 1772–1786 (2021).
42. Bourdareau, S. et al. Histone modifications during the life cycle of the brown alga Ectocarpus. *Genome Biol.* **22**, 12 (2021).
43. Gueno, J. et al. Chromatin landscape associated with sexual differentiation in a UV sex determination system. *Nucleic Acids Res.* **50**, 3307–3322 (2022).
44. Vigneau, J. et al. Interactions between U and V sex chromosomes during the life cycle of *Ectocarpus*. *Development* **151**, dev202677 (2024).
45. Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
46. Lipinska et al. Sexual dimorphism and the evolution of sex-biased gene expression in the brown alga *Ectocarpus*. *Mol. Biol. Evol.* **32**, 1581–1597 (2015).
47. Cossard, G. G. et al. Selection drives convergent gene expression changes during transitions to co-sexuality in haploid sexual systems. *Nat. Ecol. Evol.* **6**, 579–589 (2022).
48. Lipinska, A. P. et al. Multiple gene movements into and out of haploid sex chromosomes. *Genome Biol.* **18**, 104 (2017).
49. Barrera-Redondo et al. Origin and evolutionary trajectories of brown algal sex chromosomes. Preprint at *bioRxiv* https://doi.org/10.1101/2024.01.15.575685 (2024).
50. Avia, K. et al. Genetic diversity in the UV sex chromosomes of the brown alga *Ectocarpus*. *Genes* **9**, https://doi.org/10.3390/genes9060286 (2018).
51. Talbert, P. B. & Henikoff, S. What makes a centromere? *Exp. Cell Res.* **389**, 111895 (2020).
52. Yap, K. L. & Zhou, M.-M. Structure and mechanisms of lysine methylation recognition by the chromodomain in gene transcription. *Biochemistry* **50**, 1966–1980 (2011).
53. Müller, D. G., Kapp, M. & Knippers, R. Viruses in marine brown algae. *Adv. Virus Res.* **50**, 49–67 (1998).
54. Delaroque, N., Maier, I., Knippers, R. & Müller, D. G. Persistent virus integration into the genome of its algal host, Ectocarpus siliculosus (Phaeophyceae). *J. Gen. Virol.* **80**, 1367–1370 (1999).
55. Delaroque, N. et al. The complete DNA sequence of the *Ectocarpus siliculosus* virus EsV-1 genome. *Virology* **287**, 112–132 (2001).
56. Alvarez-Ponce, D. & Fares, M. A. Evolutionary rate and duplicability in the Arabidopsis thaliana protein-protein interaction network. *Genome Biol. Evol.* **4**, 1263–1274 (2012).
57. Fransz, P., De Jong, J. H., Lysak, M., Castiglione, M. R. & Schubert, I. Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl. Acad. Sci. USA* **99**, 14584–14589 (2002).
58. Doğan, E. S. & Liu, C. Three-dimensional chromatin packing and positioning of plant genomes. *Nat. Plants* **4**, 521–529 (2018).
59. Ong, C.-T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* **15**, 234–246 (2014).
60. Luthringer, R. et al. Repeated co-option of HMG-box genes for sex determination in brown algae and animals. *Science* https://doi.org/10.1126/science.adk5466 (2024).
61. Corcoran, P. et al. Introgression maintains the genetic integrity of the mating-type determining chromosome of the fungus *Neurospora tetrasperma*. *Genome Res.* **26**, 486–498 (2016).
62. Carpentier, F. et al. Convergent recombination cessation between mating-type genes and centromeres in selfing anther-smut fungi. *Genome Res.* **29**, 944–953 (2019).
63. Coelho, S. M. & Umen, J. Switching it up: algal insights into sexual transitions. *Plant Reprod.* **34**, 287–296 (2021).
64. Belton, J.-M. et al. The conformation of yeast chromosome III is mating type dependent and controlled by the recombination enhancer. *Cell Rep.* **13**, 1855–1867 (2015).
65. Li, M., Fine, R. D., Dinda, M., Bekiranov, S. & Smith, J. S. A Sir2-regulated locus control region in the recombination enhancer of Saccharomyces cerevisiae specifies chromosome III structure. *PLOS Genet.* **15**, e1008339 (2019).
66. Dubin, M., Fuchs, J., Gräf, R., Schubert, I. & Nellen, W. Dynamics of a novel centromeric histone variant CenH3 reveals the evolutionary ancestral timing of centromere biogenesis. *Nucleic Acids Res.* **38**, 7526–7537 (2010).
67. Glöckner, G. & Heidel, A. J. Centromere sequence and dynamics in Dictyostelium discoideum. *Nucleic Acids Res.* **37**, 1809–1816 (2009).
68. Fang, Y. et al. Long transposon-rich centromeres in an oomycete reveal divergence of centromere features in Stramenopila-Alveolata-Rhizaria lineages. *PLOS Genet.* **16**, e1008646 (2020).
69. Diner, R. E. et al. Diatom centromeres suggest a mechanism for nuclear DNA acquisition. *Proc. Natl. Acad. Sci.* USA **114**, https://doi.org/10.1073/pnas.1700764114 (2017).
70. Gao, X., Hou, Y., Ebina, H., Levin, H. L. & Voytas, D. F. Chromodomains direct integration of retrotransposons to heterochromatin. *Genome Res.* **18**, 359–369 (2008).
71. Neumann, P. et al. Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mob. DNA* **2**, 4 (2011).
72. Ustyantsev, K., Blinov, A. & Smyshlyaev, G. Convergence of retrotransposons in oomycetes and plants. *Mob. DNA* **8**, 4 (2017).
73. Llorens, C., Muñoz-Pomer, A., Bernad, L., Botella, H. & Moya, A. Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. *Biol. Direct* **4**, 41 (2009).
74. Knipe, D. M. et al. Snapshots: Chromatin control of viral infection. *Virology* **435**, 141–156 (2013).
75. McKeown, D. et al. Phaeoviral infections are present in Macrocystis, Ecklonia and Undaria (Laminariales) and are influenced by wave exposure in ectocarpales. *Viruses* **10**, 410 (2018).
76. Stam, M., Tark-Dame, M. & Fransz, P. 3D genome organization: a role for phase separation and loop extrusion? *Curr. Opin. Plant Biol.* **48**, 36–46 (2019).
77. Coelho, S. M. et al. How to cultivate *Ectocarpus*. *Cold Spring Harb. Protoc.* **2012**, 258–261 (2012).
78. Liu, C., Cheng, Y.-J., Wang, J.-W. & Weigel, D. Prominent topologically associated domains differentiate global chromatin packing in rice from Arabidopsis. *Nat. Plants* **3**, 742–748 (2017).
79. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one* **9**, e112963 (2014).
80. Yeo, S., Coombe, L., Warren, R. L., Chu, J. & Birol, I. ARCS: scaffolding genome drafts with linked reads. *Bioinformatics* **34**, 725–731 (2018).
81. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* **30**, 2114–2120 (2014).

82. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
83. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
84. Midekso, F. D. & Yi, G. RFfiller: a robust and fast statistical algorithm for gap filling in draft genomes. *PeerJ* **10**, e14186 (2022).
85. Xu, M. et al. TGS-GapCloser: A fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *GigaScience* **9**, giaa094 (2020).
86. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
87. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
88. Laetsch, D. & Blaxter, M. BlobTools: Interrogation of genome assemblies. *F1000Research* https://doi.org/10.12688/f1000research.12232.1 (2017).
89. Durand, N. C. et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst.* **3**, 95–98 (2016).
90. Langmead, B. & Salzberg, S. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
91. Lee, S., Bakker, C. R., Vitzthum, C., Alver, B. H. & Park, P. J. Pairs and Pairix: a file format and a tool for efficient storage and retrieval for Hi-C read pairs. *Bioinformatics* **38**, 1729–1731 (2022).
92. Open2C, et al. *Cooltools*: Enabling high-resolution Hi-C analysis in Python. *PLoS Comput. Biol.* **20**, e1012067 (2024).
93. Ewels, P. A. et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).
94. Krueger, F. Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Babraham Institute (2015).
95. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
96. Gaspar, J. M. Improved peak-calling with MACS2. bioRxiv (2018).
97. Wu, T. D., Reeder, J., Lawrence, M., Becker, G. & Brauer, M. J. GMAP and GSNAP for genomic sequence alignment: Enhancements to speed, accuracy, and functionality. *Methods Mol. Biol.* **1418**, 283–334 (2016).
98. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
99. Love, B., Sidebotham, M., Fenwick, J., Harvey, S. & Fairbrother, G. "Unscrambling what's in your head": A mixed method evaluation of clinical supervision for midwives. *Women Birth* **30**, 271–281 (2017).
100. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA* **117**, 9451–9457 (2020).
101. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids. Res.* **27**, 573–580 (1999).
102. Goubert, C. et al. A beginner's guide to manual curation of transposable elements. *Mobile DNA* **13**, 7 (2022).
103. Kohany, O., Gentles, A. J., Hankus, L. & Jurka, J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinforma.* **7**, 474 (2006).
104. Hoencamp, C. et al. 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. *Science* **372**, 984–989 (2021).
105. Oluwadare, O., Zhang, Y. & Cheng, J. A maximum likelihood algorithm for reconstructing 3D structures of human chromosomes from chromosomal contact data. *BMC Genomics* **19**, 161 (2018).


## Acknowledgements

This work was supported by the MPG, the ERC (grant n. 864038), the Moore Foundation (GBMF11489), and the Bettencourt-Schuller Foundation. R.C. is supported by a grant HORIZON-MSCA-2022-PF-01 (Project ID: 101109906). We thank the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) (031A532B, 031A533A, 031A533B, 031A534A, 031A535A, 031A537A, 031A537B, 031A537C, 031A537D, 031A538A). We thank Remy Luthringer and Andrea Belkacemi for assistance with the algal cultures.

## Author contributions

P.L.: Investigation (lead), Formal analysis (lead), Visualization (lead), Writing – original draft (equal). J.V.: Investigation (equal), Methodology (supporting), Formal analysis (supporting). R.C.: Investigation (equal), Methodology (supporting), Visualization (equal), Formal analysis (equal), Writing – review and editing (supporting). J.B.R. Investigation (equal), Methodology (supporting), Visualization (equal), Formal analysis (equal). E.A. and C.M.: Investigation (supporting). M.B.:Methodology (supporting). F.B.H. and C.L.: Data curation (equal), Visualization (supporting), Formal analysis (equal), supervision (equal), Writing – review and editing (supporting). S.M.C.: Conceptualization (lead), Funding acquisition (lead), Methodology (equal), Project administration (lead), Supervision (lead), Visualization (supporting), Writing – original draft (equal), Writing – review and editing (lead).

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-53453-5.

**Correspondence** and requests for materials should be addressed to Susana M. Coelho.

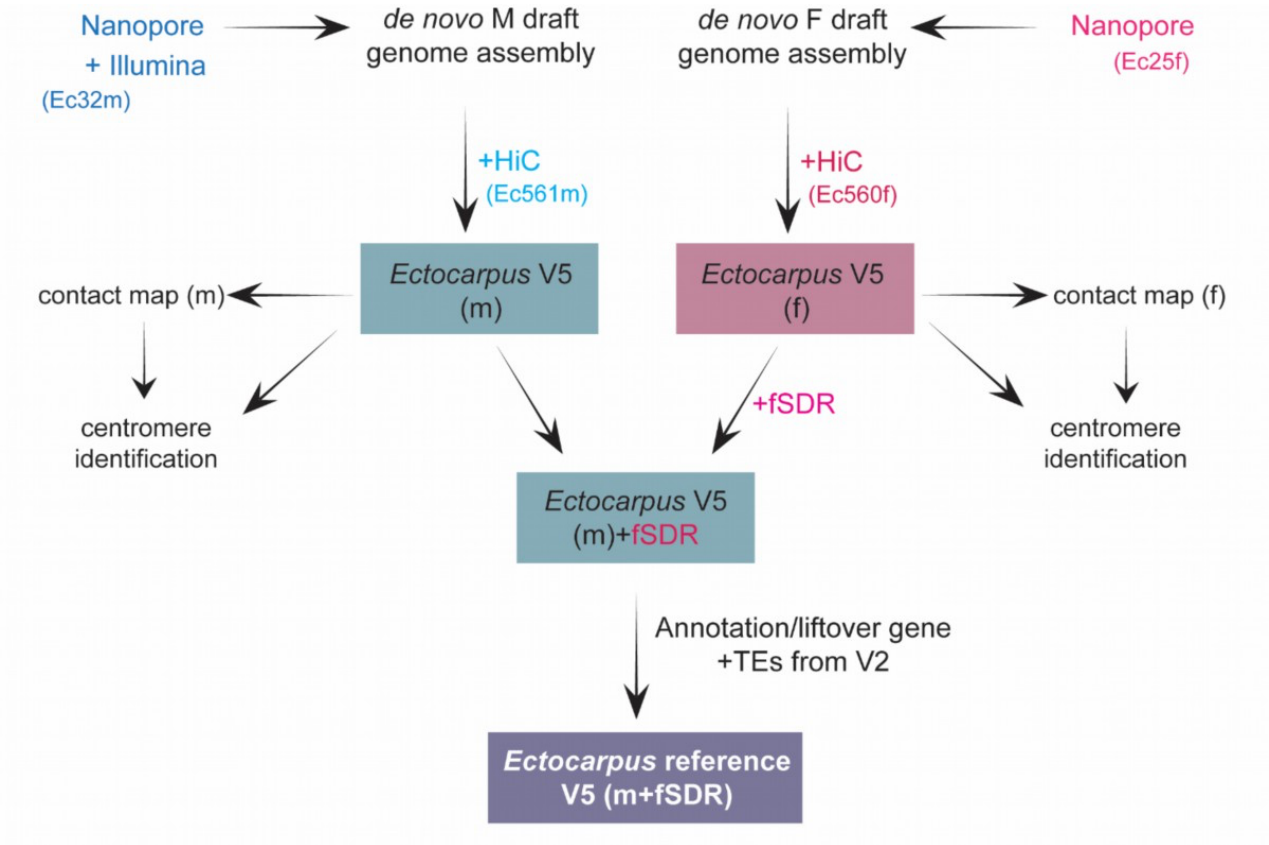
**Peer review information** *Nature Communications* thanks Célia Baroux and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

# Supplementary Information

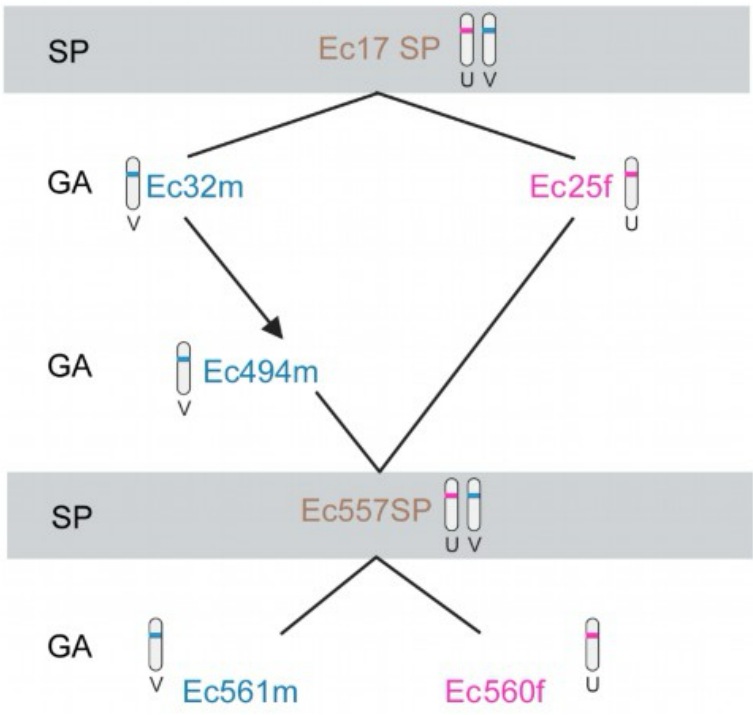
3D chromatin maps of a brown alga reveal U/V sex chromosome spatial organization

Liu et al.

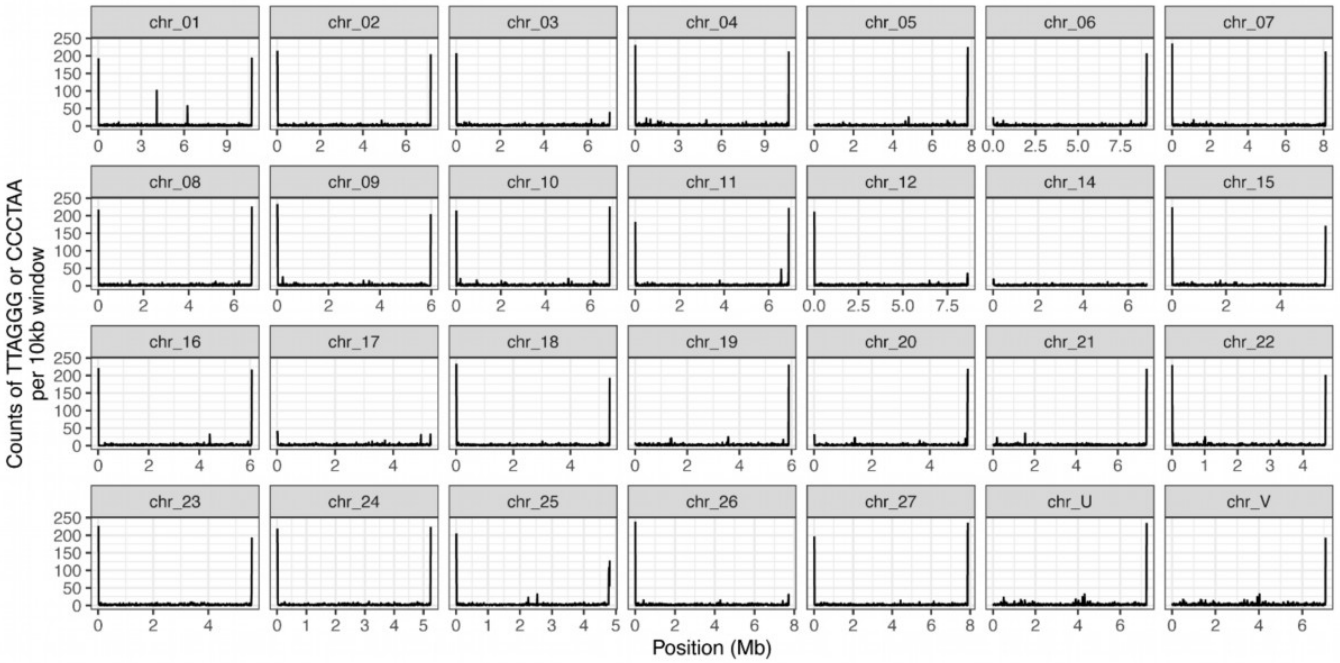
# Table of Contents



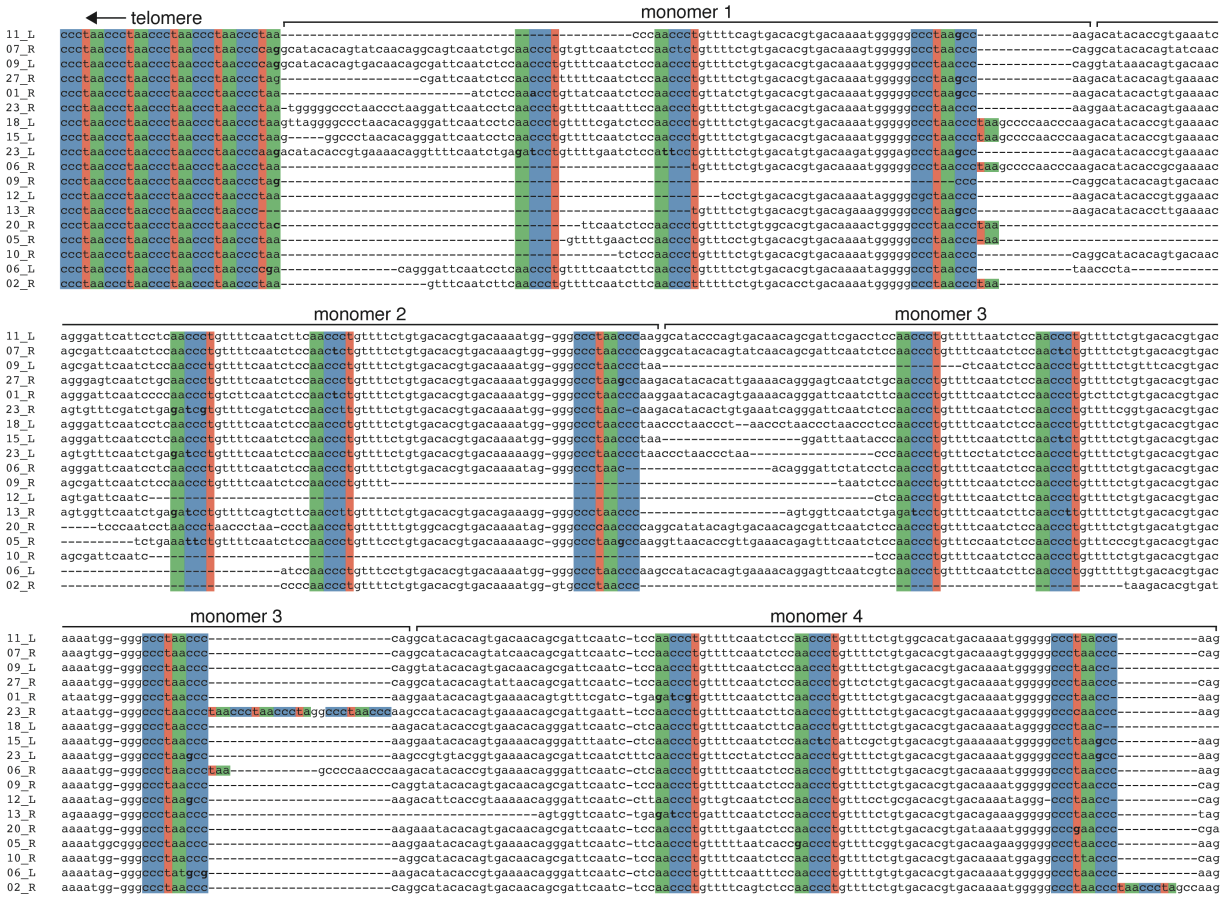

**Supplementary Fig. 1: Schematic view of the approach used to reach a high-quality genome assembly of male and female *Ectocarpus***. De novo draft genomes from male and female siblings (Ec32 male and Ec25 female) were generated using Nanopore, and the male genome was polished using illumine reads. Genomes were further assembled using Hi-C data from Ec560 and Ec561 (near isogenic male and female lines[1]. This resulted in the generation of male and female V5 genomes that were used for producing high resolution male and female contact maps. In order to have only one 'reference' genome, we chose to use the male V5 reference genome and complemented it with the female-specific sex-determining contig (SDR). Therefore, the final 'reference' *Ectocarpus* genome V5 is composed of high-quality male genome that includes both the male and the female SDR. Annotation of this V5 reference genome was performed by lifting over gene and TEs annotations from the *Ectocarpus* V2 genome[2]. Centromeres were identified based on the contact maps and their annotation further refined (see methods for details). *Ectocarpus* strain numbers are given inside brackets (see also **Supplementary Table** 1).



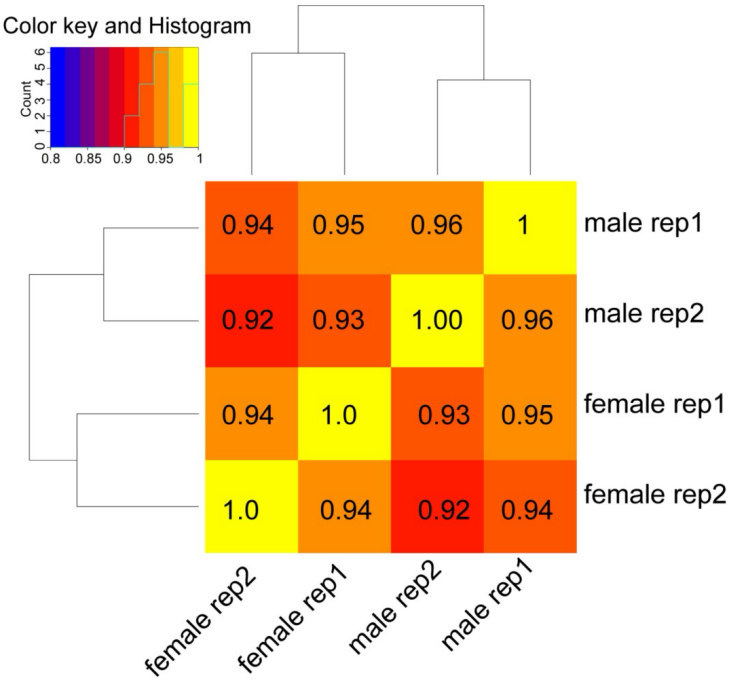

**Supplementary Fig. 2: Pedigree of the *Ectocarpus* strains used in this study.** *SP, diploid sporophyte; GA, gametophyte; m, male gametophyte; f, female gametophyte.*



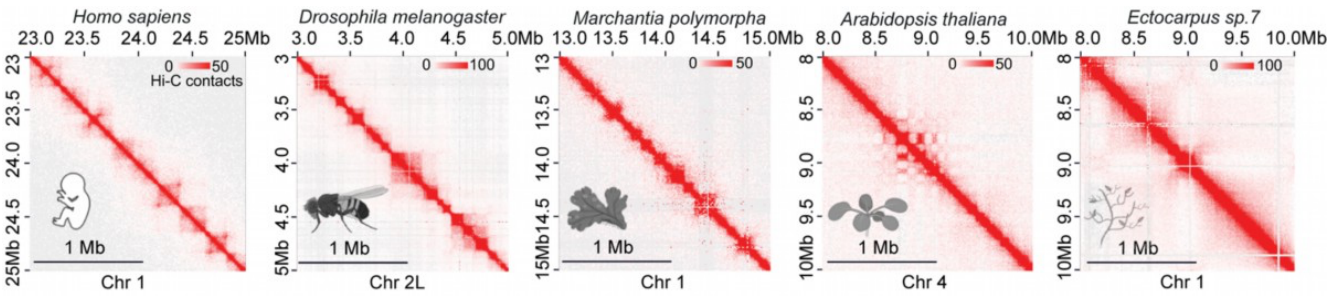

**Supplementary Fig. 3: Distribution of telomere repeat motif (TTAGGG or CCCTAA) in the *Ectocarpus* haploid genome.**



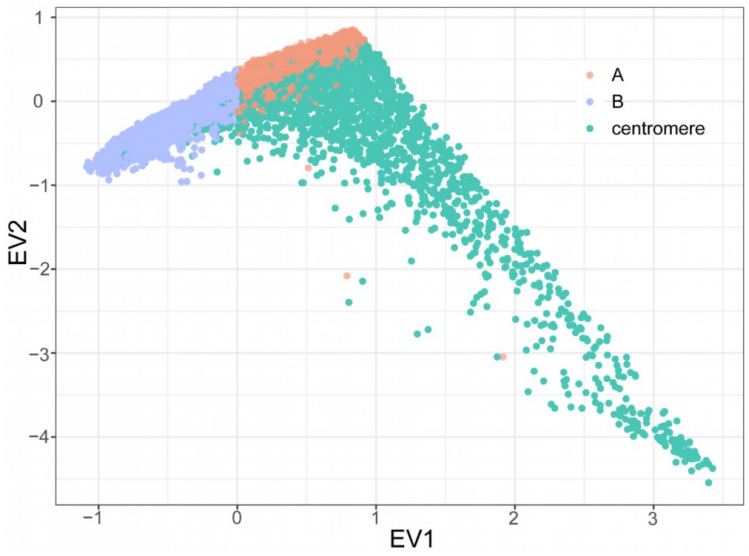

**Supplementary Fig. 4: The organization of *Ectocarpus* subtelomeres**. Alignment of 18 subtelomeres (e.g. 11_L is the left extremity of chromosome 11) showing the transition from the telomere to the subtelomeric satellite. The first four monomers of the ~98 bp satellite are shown. Telomeres and telomeric motifs present in the subtelomeric satellite are colored.



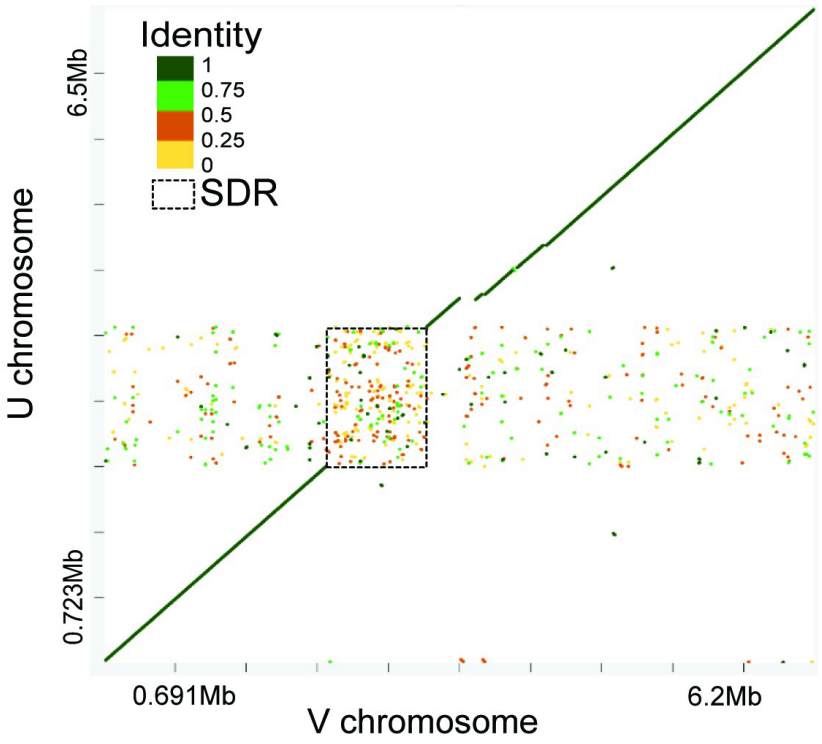

**Supplementary Fig. 5: Quality control of biological replicates of Hi-C data**. Pearson correction of biological replicates of *Ectocarpus* male and female Hi-C data at 10k bin size**.**



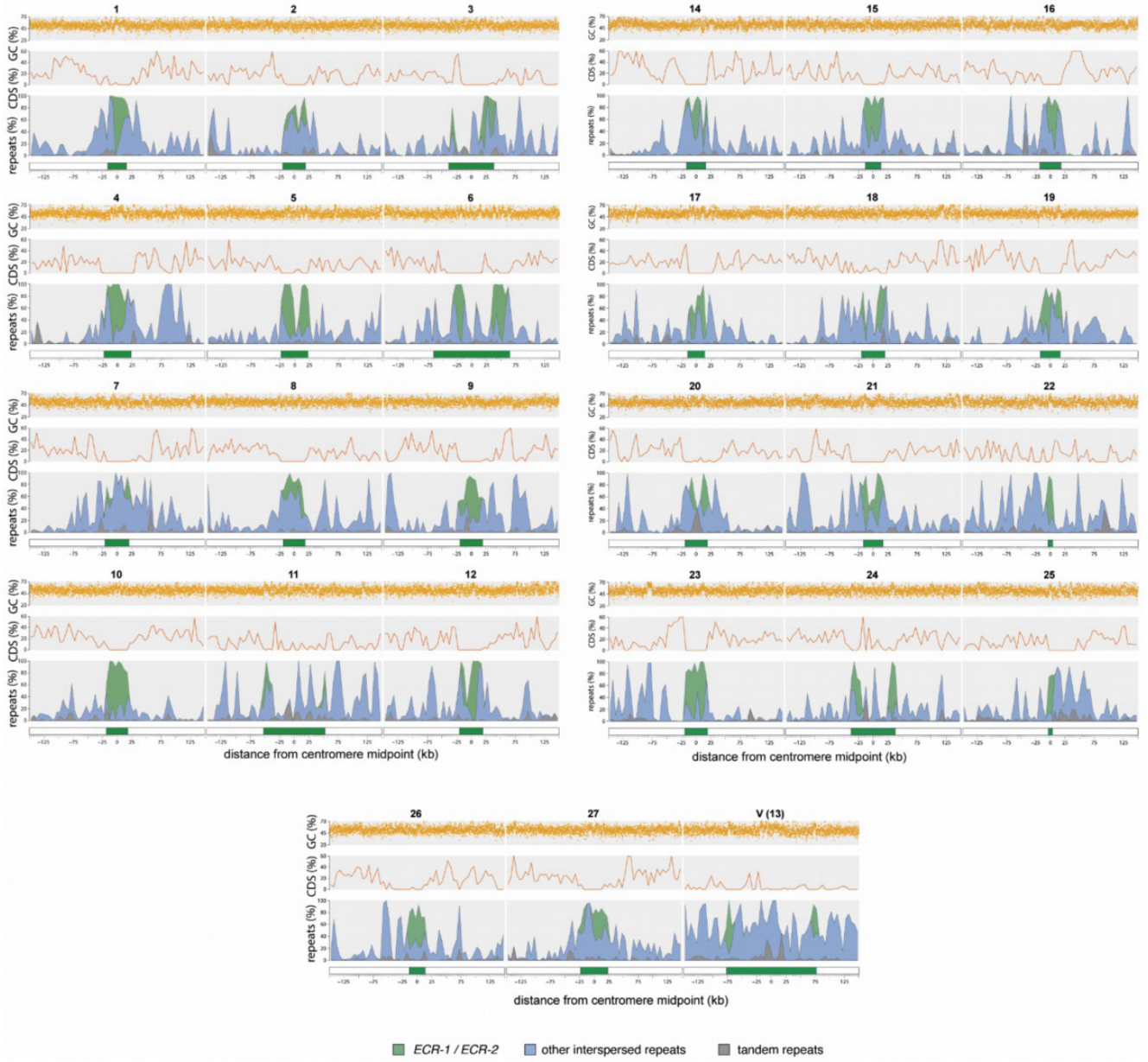

**Supplementary Fig. 6: No prominent TAD patterns are observed in *Ectocarpus*.** Examples of Hi-C maps from different species representing TADs patterns in *H. sapiens* (chromosome 1,) *Drosophila* (chromosome 2L,[3]), *Marchantia* (chromosome 1,[4]), *Arabidopsis* (chromosome 4,[5]) and *Ectocarpus* (chromosome 4, our paper). HiC maps of the different species were obtained using Juicerbox[6] at 10kb resolution. Note that TADs are not an obvious feature of *Arabidopsis* nor *Ectocarpus* genomes.



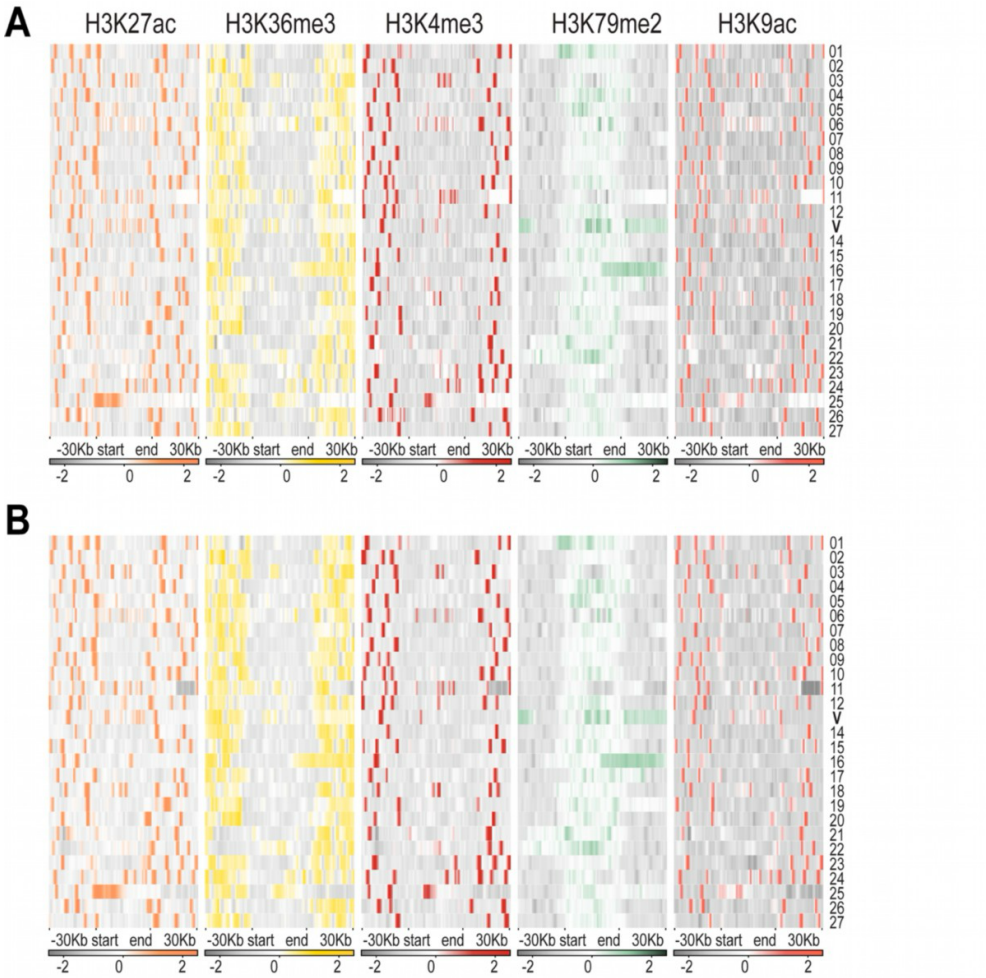

**Supplementary Fig. 7: Centromeres form distinct sub-compartments.** Centromere is represented as a sub-compartment, which could be separated from compartments by E1(PC1) and E2(PC2).



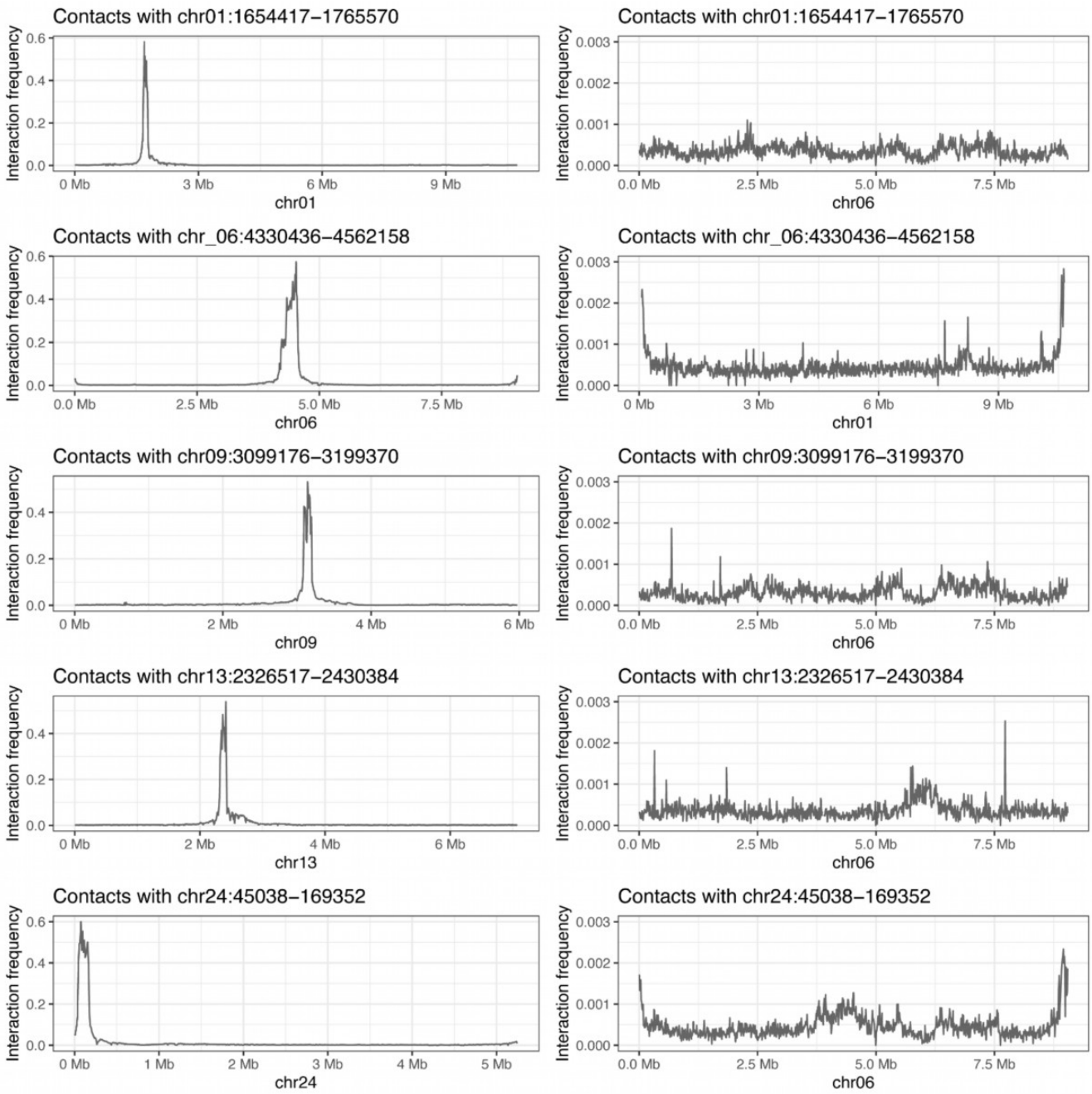

**Supplementary Fig. 8: Sequence alignment between the U and V sex chromosomes**. The matches are presented as colored lines. The colors correspond to identity values that have been clustered in four groups (below 25%, between 25% and 50%, between 50% and 75% and over 75%), the dash box shows SDRs.




**Supplementary Fig. 9: Sequence characteristics of *Ectocarpus* centromeres.** Putative centromeres and flanking regions for all chromosomes from the male Ec32 V5 assembly. The centromere (green box) is defined as the region from the first to the last copy of *ECR* elements. The repeats panel is shown as a stacked area plot, and the percentage of each repeat type is plotted in 5 kb windows. Coding sequence (CDS) density is plotted in 5 kb windows, and GC content is plotted in 100 bp windows.




**Supplementary Fig. 10: Heatmaps of histone marks around centromeres.** For each heatmap, the log2(ChIP/input) is plotted over the putative centromeres and their 30Kb surrounding regions using a bin size of 100bp using uniquely mapped (A) and multi-mapped reads (B).




**Supplementary Fig. 11: Examples of virtual 4C-like plot of H3K79me2 domains larger than 100 kb.** Intra/inter chromosomal interaction frequency of H3Km79me2 domains.

## Supplementary References


1. Gueno, J. *et al.* Chromatin landscape associated with sexual differentiation in a UV sex determination system. *Nucleic Acids Res* **50**, 3307–3322 (2022).
2. Cormier, A. *et al.* Re-annotation, improved large-scale assembly and establishment of a catalogue of noncoding loci for the genome of the model brown alga Ectocarpus. *New Phytologist* **214**, (2017).
3. Ray, J. *et al.* Chromatin conformation remains stable upon extensive transcriptional changes driven by heat shock. *Proc Natl Acad Sci U S A* **116**, (2019).
4. Karaaslan, E. S. *et al.* Marchantia TCP transcription factor activity correlates with three-dimensional chromatin structure. *Nat Plants* **6**, 1250–1261 (2020).
5. Yin, X. *et al.* Binding by the Polycomb complex component BMI1 and H2A monoubiquitination shape local and long-range interactions in the Arabidopsis genome. *Plant Cell* (2023) doi:10.1093/plcell/koad112.
6. Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* **3**, 99–101 (2016).

nature portfolio

Corresponding author(s): Susana Coelho

Last updated by author(s): Sep 30, 2024

# Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☒ | ☐ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☒ | ☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☒ | ☐ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | No software was used for data collection |
| Data analysis | Tools for reads analysis and genome assembly: ONT Guppy, Canu assembler and 3D-DNA pipeline, TGS-GapCloser, BUSCO;<br>Tools for Hi-C analysis: juicer, juicerbox, pairix, cooler, cooltools, HiCdat, HiContacts;<br>Tools for RNA-seq analysis: GSNAP aligner, featureCount, DEseq2;<br>Tools for ChIP-seq (re-analysis): Trim Galore, nf-core ChIP-seq pipeline, BWA, MACS2,deeptools, IGV, ggChIPvis;<br>Tools for repeats analysis: RepeatModeler, RepeatMasker, Tandem Repeats Finder, Repbase Censor;<br>Parameters are described in relative sections in "Methods" |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The Nanopore and Hi-C data generated in this study have been deposited in the NCBI database under the project number PRJNA1105946. Ectocarpus V5 genome and gene annotation, the processed Hi-C, ChIP-seq and RNA-seq data are available in Edmond of Max Planck Digital Library collection (doi.org/10.17617/3.QXUAMN and doi.org/10.17617/3.NQDSLW). The RNA-seq and ChIP-seq datasets used in this study were retrieved from the NCBI Gene Expression Omnibus repository: PRJNA1055718.

## Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

| | |
|---|---|
| Reporting on sex and gender | *Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used.*<br>*Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.*<br>*Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.* |
| Reporting on race, ethnicity, or other socially relevant groupings | *Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).*<br>*Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)*<br>*Please provide details about how you controlled for confounding variables in your analyses.* |
| Population characteristics | *Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."* |
| Recruitment | *Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.* |
| Ethics oversight | *Identify the organization(s) that approved the study protocol.* |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | n.a |
| Data exclusions | no data was excluded |
| Replication | two replicates were used for the HiC experiments, following correlation analysis these two replicates were pooled together (as usual fort his type of analysis) |
| Randomization | algal samples were grown in a random way in the culture chambers |
| Blinding | n.a. |

# Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Study description | *Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).* |
| Research sample | *State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.* |
| Sampling strategy | *Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.* |
| Data collection | *Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.* |
| Timing | *Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.* |
| Data exclusions | *If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.* |
| Non-participation | *State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.* |
| Randomization | *If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.* |

# Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Study description | *Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.* |
| Research sample | *Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.* |
| Sampling strategy | *Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.* |
| Data collection | *Describe the data collection procedure, including who recorded the data and how.* |
| Timing and spatial scale | *Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken* |
| Data exclusions | *If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.* |
| Reproducibility | *Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.* |
| Randomization | *Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.* |
| Blinding | *Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.* |

Did the study involve field work? ☐ Yes ☒ No

## Field work, collection and transport

| | |
|---|---|
| Field conditions | *Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).* |
| Location | *State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).* |
| Access & import/export | *Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).* |
| Disturbance | *Describe any disturbance caused by the study and how it was minimized.* |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |
| ☒ | ☐ Plants |

### Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Antibodies

| | |
|---|---|
| Antibodies used | *Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.* |
| Validation | *Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.* |

## Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

| | |
|---|---|
| Cell line source(s) | *State the source of each cell line used and the sex of all primary cell lines and cells derived from human participants or vertebrate models.* |
| Authentication | *Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.* |
| Mycoplasma contamination | *Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.* |
| Commonly misidentified lines (See ICLAC register) | *Name any commonly misidentified cell lines used in the study and provide a rationale for their use.* |

## Palaeontology and Archaeology

| | |
|---|---|
| Specimen provenance | *Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.* |
| Specimen deposition | *Indicate where the specimens have been deposited to permit free access by other researchers.* |

Dating methods

*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight

*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals

*For laboratory animals, report species, strain and age OR state that the study did not involve laboratory animals.*

Wild animals

*Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*

Reporting on sex

*Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify reasons for lack of sex-based analysis.*

Field-collected samples

*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*

Ethics oversight

*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration

*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*

Study protocol

*Note where the full trial protocol can be accessed OR if not available, explain why.*

Data collection

*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*

Outcomes

*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

## Dual use research of concern

Policy information about dual use research of concern

### Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

| No | Yes | |
|---|---|---|
| ☒ | ☐ | Public health |
| ☒ | ☐ | National security |
| ☒ | ☐ | Crops and/or livestock |
| ☒ | ☐ | Ecosystems |
| ☒ | ☐ | Any other significant area |

## Experiments of concern

Does the work involve any of these experiments of concern:

| No | Yes | |
|---|---|---|
| ☒ | ☐ | Demonstrate how to render a vaccine ineffective |
| ☒ | ☐ | Confer resistance to therapeutically useful antibiotics or antiviral agents |
| ☒ | ☐ | Enhance the virulence of a pathogen or render a nonpathogen virulent |
| ☒ | ☐ | Increase transmissibility of a pathogen |
| ☒ | ☐ | Alter the host range of a pathogen |
| ☒ | ☐ | Enable evasion of diagnostic/detection modalities |
| ☒ | ☐ | Enable the weaponization of a biological agent or toxin |
| ☒ | ☐ | Any other potentially harmful combination of experiments and agents |

# Plants

Seed stocks

*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*

Novel plant genotypes

*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*

Authentication

*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

# Flow Cytometry

## Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

Sample preparation

*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*

Instrument

*Identify the instrument used for data collection, specifying make and model number.*

Software

*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*

Cell population abundance

*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*

Gating strategy

*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

# Magnetic resonance imaging

## Experimental design

Design type

*Indicate task or resting state; event-related or block design.*

Design specifications

*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*

Behavioral performance measures

*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

## Acquisition

Imaging type(s)

*Specify: functional, structural, diffusion, perfusion.*

Field strength

*Specify in Tesla*

Sequence & imaging parameters

*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*

Area of acquisition

*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

## Preprocessing

Preprocessing software

*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*

Normalization

*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*

Normalization template

*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*

Noise and artifact removal

*Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring

*Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

## Statistical modeling & inference

Model type and settings

*Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*

Effect(s) tested

*Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference

*Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*

(See Eklund et al. 2016)

Correction

*Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

## Models & analysis

| n/a | Involved in the study |
|---|---|
| ☐ | ☐ Functional and/or effective connectivity |
| ☐ | ☐ Graph analysis |
| ☐ | ☐ Multivariate modeling or predictive analysis |

Functional and/or effective connectivity

*Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*

Graph analysis

*Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*

Multivariate modeling and predictive analysis

*Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*

nature ecology & evolution

Article

https://doi.org/10.1038/s41559-025-02838-w

# Origin and evolutionary trajectories of brown algal sex chromosomes

Received: 22 September 2024

Accepted: 22 July 2025

Published online: 25 August 2025

Check for updates

Josué Barrera-Redondo[1,9,11], Agnieszka P. Lipinska[1,2,11], Pengfei Liu[1], Erica Dinatale[1], Guillaume Cossard[1], Kenny Bogaert[1], Masakazu Hoshino[1,10], Rory J. Craig[1], Komlan Avia[3], Goncalo Leiria[1], Elena Avdievich[1], Daniel Liesner[1], Rémy Luthringer[1], Olivier Godfroy[2], Svenja Heesch[2], Zofia Nehr[2], Loraine Brillet-Guéguen[2,4], Akira F. Peters[5], Galice Hoarau[6], Gareth Pearson[7], Jean-Marc Aury[8], Patrick Wincker[8], France Denoeud[8,12], J. Mark Cock[2,12], Fabian B. Haas[1,12] & Susana M. Coelho[1,12]

Research on the biology and evolution of sex chromosomes has primarily focused on diploid XX/XY and ZW/ZZ systems. In contrast, the rise, evolution and demise of U/V systems has remained an enigma. Here we analyse genomes of nine brown algal species with different sexual systems to determine the history of their sex determination. U/V sex chromosomes emerged between 450 and 224 million years ago, when a region containing the pivotal male-determinant *MIN* ceased recombining. Seven ancestral genes within the sex-determining region show remarkable conservation over this vast evolutionary time, although nested inversions caused expansions of the sex locus, independently in each lineage. We evaluate whether these expansions are associated with increased morphological complexity and sexual differentiation, and show that taxonomically restricted genes evolve unexpectedly often in U and V chromosomes. We also investigate two situations in which U/V-linked regions have changed. First, we demonstrate that convergent evolution of two monoicous species occurred by ancestral males acquiring U-specific genes. Second, the *Fucus* dioecious system involves new sex-determining gene(s), acting upstream of formerly V-specific genes during development. Both situations have led to the demise of U and V chromosomes and erosion of their specific genomic characteristics.

The mechanisms controlling the development of male or female identities, or co-sexuality, when individuals express both sex functions, vary widely across different organisms[1,2]. In species with separate sexes, sex chromosomes may be present, carrying a sex-determining region (SDR)[3] that encodes factors directing sex identity and which often does not undergo recombination in the heterogametic sex (XY or ZW)[4] of diploid species (dioecious), or in the diploid stage of haploid-dominant (dioicous) species. Sex chromosomes have independently evolved from autosomes multiple times and may be subject to specific evolutionary forces, including differential selection between sexes, asymmetrical expression of deleterious mutations and hemizygosity, meiotic silencing and dosage compensation[3].

Research on the biology and evolution of sex chromosomes has primarily focused on diploid XX/XY and ZW/ZZ systems in mammals, birds, fish, *Drosophila* and diploid plants[4,5]. U/V haploid sex-determination systems, such as those of bryophytes and algae[6,7], have been less explored. In U/V systems, sex is not determined at fertilization but during meiosis, when haploid spores inherit either a

A full list of affiliations appears at the end of the paper. e-mail: susana.coelho@tuebingen.mpg.de

U chromosome, and will develop into a female gametophyte, or a V chromosome, controlling male gametophyte formation[8]. These fundamental inheritance differences between U/V and XX/XY or ZW/ZZ systems have broad evolutionary and genomic implications[9,10]. However, so far, only the U/V systems of the brown alga *Ectocarpus* and the U/Vs of four distantly related bryophyte taxa[11–14] have been fully sequenced and assembled into chromosomes. While these studies helped understand the genomic structure of bryophyte U- and V-linked regions, the species involved diverged ~500 Ma (million years ago) and do not share homologous U/V chromosomes[15]. As a result, we still lack a broad comparative view across multiple homologous U/V systems that would inform a reconstruction of their evolutionary history. Brown algae represent exceptional models for studying sex chromosome evolution because they display diverse reproductive systems, life cycles and sex chromosome systems in a single lineage[16]. Their ancestral state probably involved separate sexes[16], suggesting that their sex chromosomes could share a common origin. Here we study the origin, evolution and demise of U/V sex chromosomes in the brown algae.

## Results

### The origin of brown algal sex chromosomes

We focused on species covering the phylogenetic, morphological and reproductive diversity of the brown algal clade[17] and their closest extant outgroup, *Schizocladia ischiensis*[17,18]. We substantially improved the brown algal genome datasets available[18–20] to reach chromosome or near-chromosome-level genome assemblies (Extended Data Table 1 and Supplementary Table 1). This revealed that brown algae have 27–33 chromosomes and largely conserved macrosynteny (Fig. 1a).

We identified the female (U) and male (V) sex-determining regions (SDRs) in the dioicous species using a combination of bioinformatic and experimental approaches (see 'Discovery of the U/V sex determination regions' in Methods; Supplementary Figs. 1–5). All U/V species share the same, albeit highly rearranged, ancestral sex chromosome, showing remarkable stability despite the large evolutionary time (Fig. 1b,c and Extended Data Table 1). The recombination suppression event leading to the birth of U/V sex chromosomes occurred after the split of *S. ischiensis* and *Dictyota dichotoma*, ~450–224 Ma[21] (Fig. 1a). The male-determining gene *MIN*[22] is the only V-specific gene consistently present in all V-SDRs of the dioicous species. We note that one dioecious (*Fucus serratus*) and two monoicous (haploid, co-sexual *Chordaria linearis* and *Desmarestia dudresnayi*) species lack U/V sex chromosomes but still retain *MIN* on a chromosome homologous to the ancestral U/V ('U/V-homologue' hereafter). The outgroup *S. ischiensis* has low synteny with the brown algae and exhibits putative fusion-with-mixing events[23] (Fig. 1a and Extended Data Fig. 1).

We next examined the U/V-SDRs by comparing male and female genome assemblies (Methods). The SDRs contain a small number of genes overall (between 18 and 52), and compared with the pseudo-autosomal regions (PARs) (between 229 and 904), with considerable variation in gene content and size across species, the smallest being found in the Ectocarpales (*Ectocarpus* sp. 7, *Ectocarpus crouaniorum*, *Scytosiphon promiscuus*; Fig. 1b,c, Extended Data Fig. 2a and Extended Data Table 1). SDR size differences across species are strongly correlated with the number of genes ($R^2$ = 0.97; Extended Data Fig. 2b) and the repeat content ($R^2$ = 0.99; Extended Data Fig. 2c) inside these regions. Many genes located in the PARs of the Ectocarpales are within the V-SDRs of *Undaria pinnatifida*, *Desmarestia herbacea* and *D. dichotoma*, indicating that the SDR boundaries have changed across species. The boundary differences coincide with extensive structural rearrangements, particularly inversions, even among closely related taxa (Fig. 1b,c and Extended Data Fig. 3). Note that the centromere in the V chromosome of *Ectocarpus* is found within the SDR[19], so we cannot exclude that a centromere-related suppression of recombination may have preceded the inversion events found on the SDR[24].

Together, our results indicate that the brown algal U/V sex chromosomes evolved between 450–224 Ma, via suppressed recombination in a genomic region that contained *MIN* (henceforth male-determining locus). The presence of *MIN* in distantly related lineages could push the age of the U/V chromosomes further back in time, but more evidence would be required to establish that dioicy existed in these organisms.

### The evolution of the SDRs involved boundary expansions and gene gains

The brown algal U- and V-SDRs carry homologous genes (gametologue pairs), indicating descent from a common ancestral region (Supplementary Table 2). *Ectocarpus* sp. 7 and *D. herbacea* show similar ratios of gametologues and U- or V-specific genes (16/14 and 11/7 gametologue/sex-specific genes in *Ectocarpus* and *D. herbacea*, respectively; Supplementary Table 2). Only ten genes share SDR orthologues between both species, while the rest were mostly acquired independently in the SDR of each species, with one gene that was retained as a gametologue pair in *Ectocarpus* sp. 7 but lost both copies in *D. herbacea* (Supplementary Table 2). Five gametologue pairs conserved both copies in the two species, while another three gametologue pairs lost either the male or the female copy in *D. herbacea* (Supplementary Table 2). In addition, *MIN* and a U-specific gene are also conserved between species (Supplementary Table 2). Although the total number of U/V-SDR genes differs between *Ectocarpus* sp. 7 (18 genes) and *D. herbacea* (30 genes), each species shows an equal number of gametologues and sex-specific genes in its U- and V-SDRs (Supplementary Table 2). This intraspecies symmetry supports the idea that the U and V chromosomes may have undergone parallel evolutionary changes within each lineage[10,25,26]. The V-SDR of *D. herbacea* contains 20 additional genes that belong to endogenous viral elements, which are common across brown algal genomes[18].

Diploid sex chromosome in animals and plants exhibit evolutionary strata representing different recombination suppression events over time. Strata are identified by analysing synonymous substitutions ($K_s$) between male/female gametologue pairs[27] whose locations in fully X or Z-linked regions are known. However, detecting evolutionary strata in U/V systems is difficult because neither of these fully sex-linked regions recombines and gene movements and chromosome rearrangements disrupt collinearity of both chromosomes between species[25,28,29]. Moreover, in the absence of a recombining outgroup (which does not exist in brown algae), the ancestral gene order cannot be reliably inferred. In both *Ectocarpus* sp. 7 and *D. herbacea*, the V- and U-SDR rearrangements differ by inversions (Fig. 2a,b and Extended Data Fig. 4), consistent with the idea that inversions may lead to suppressed recombination between sex chromosomes. An analysis of gametologue pair divergence revealed saturated levels of $K_s$ values (Fig. 2b,c and Supplementary Table 3), further limiting the inference of evolutionary strata across brown algal SDRs. Nonetheless, the gametologue $K_s$ values are broadly consistent between orthologues in *Ectocarpus* sp. *7* and *D. herbacea*, where shared SDR gametologues between species have higher $K_s$ values and probably spent more evolutionary time diverging than the gametologues that are not shared between species (Supplementary Table 3). Furthermore, the location of gametologues with the lowest $K_s$ values in the U-SDR of *D. herbacea*, relative to the PAR genes in *Ectocarpus* sp. 7, suggests that inversions involving the entire U-SDR and adjacent PAR segments probably contributed to the expansion of the U/V-SDR boundaries in *D. herbacea*, in a process we term 'engulfment' (Extended Data Fig. 4). The expansion of the SDR boundaries in *D. herbacea* led to the engulfment of a region containing four genes in the PAR1 of *Ectocarpus* sp. 7, and a second region with 13 genes located on the PAR2 (Fig. 2d, Supplementary Table 4 and Extended Data Fig. 4). Twelve of these engulfed genes into the SDR of *D. herbacea* were retained as gametologues. These observations support a scenario where expansions in the SDR boundaries of brown algae

https://doi.org/10.1038/s41559-025-02838-w

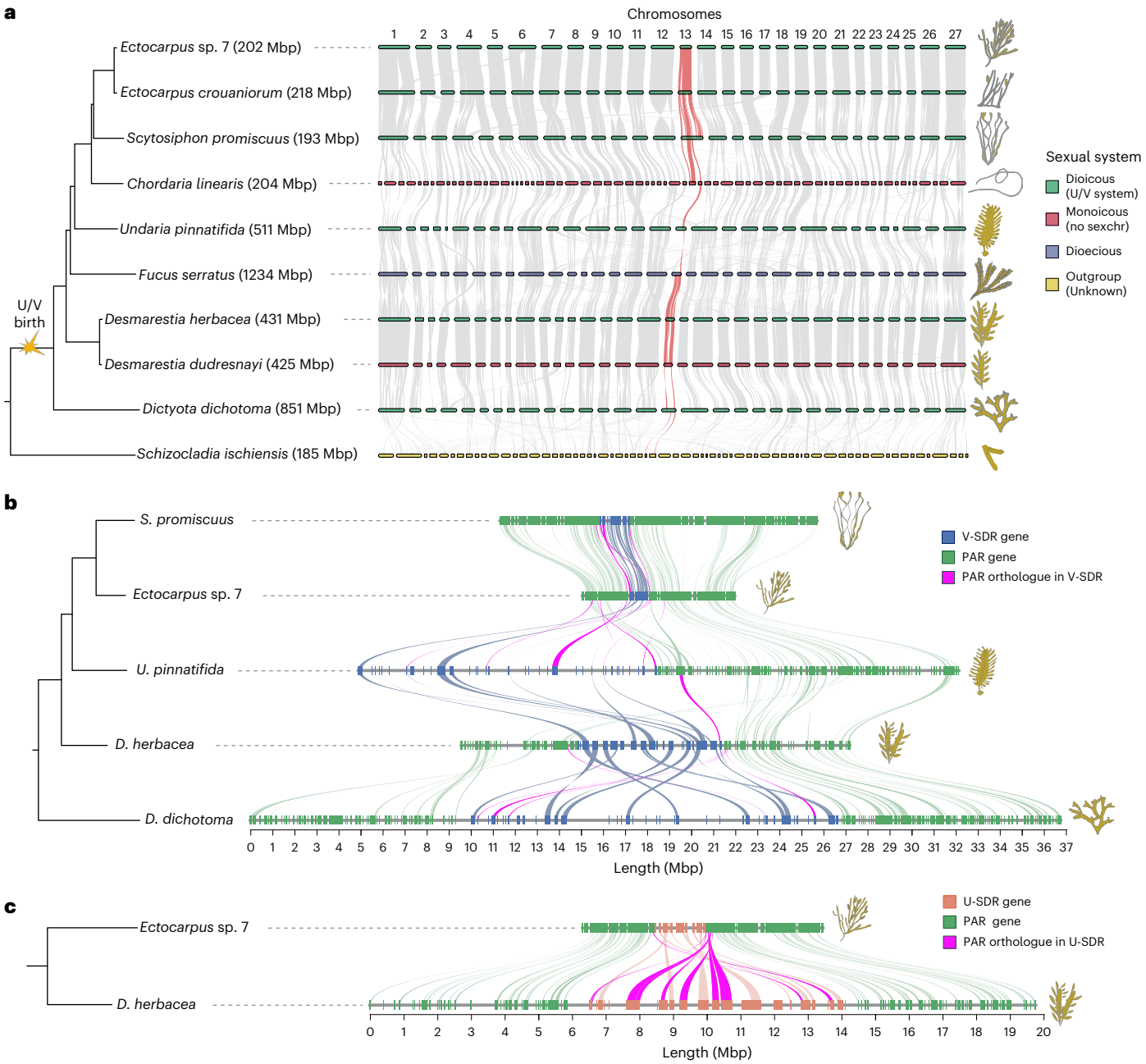



**Fig. 1 | Origins of U/V sex chromosomes in brown algae. a**, Macrosynteny plot comparing genomes of six dioicous (green), two monoicous (red), one dioecious (blue) and one outgroup species (yellow). The chromosomes were originally numbered by their physical size in the *Ectocarpus* v2 genome[141]. Note that the dioecious species *F. serratus* has a fully diploid life cycle (without gametophytes[7]). Syntenic blocks of the V sex chromosome are highlighted in red, with the emergence of U/V chromosomes shown in the phylogeny. Genome sizes are indicated in brackets. **b**, Microsynteny plot of V chromosomes in five dioicous species, highlighting the male sex-determining regions (blue) and the PARs (green). The PAR genes whose orthologues are found within the SDR of other species are highlighted in purple. **c**, Microsynteny plot of U chromosomes in two dioicous species, highlighting the female sex-determining regions (peach) and the PARs (green). The PAR genes whose orthologues are found within the SDR of other species are highlighted in purple. Note that the genome assemblies for *C. linearis* and *S. ischiensis* are not chromosome level, leading to a high number of contigs.

occur through nested inversions. Two chromatin-related transcription factors in the *Ectocarpus* PARs were independently incorporated into the SDRs of four other dioicous species (Supplementary Tables 4 and 5).

The observation of greater V-SDR gene content in early diverging lineages (such as *D. dichotoma*) than in the later-diverging Ectocarpales (Extended Data Table 1) could reflect either gene loss in the V-SDRs of Ectocarpales or independent gene gains in the V-SDRs of each lineage (as predicted in ref. 10), from an ancestral state with low V-SDR gene content that is retained in Ectocarpales. To distinguish between these possibilities, we reconstructed the ancestral SDR gene content (Supplementary Table 5), focusing on the V chromosome, as the genomic data are of better quality (Supplementary Table 1), and assuming parallel U/V-SDR evolution[10,25,26] as seen in *Ectocarpus* sp. 7 and *D. herbacea* (Fig. 2d and Supplementary Table 2). This analysis revealed that brown algal V-SDR evolution occurred via lineage-specific gene gains rather than gene loss in the Ectocarpales (Fig. 2e,f). Gene gains were caused by a combination of three processes: expansions of the SDR boundaries into the PARs, translocation of autosomal genes into the SDR and lineage-specific gene birth events within the SDR (Fig. 2e and Supplementary Table 5). Consistently, ancestral

V-SDR genes were associated with higher gametologue $K_s$ values, while independently acquired gametologues in *D. herbacea* had lower $K_s$ values (Supplementary Table 2 and Fig. 2c).

The seven genes in the ancestral V-SDR (Fig. 2f and Extended Data Fig. 2d) include the male-determinant *MIN*[22] and six V gametologues of genes that are also carried on the U chromosome (Fig. 2f and Supplementary Table 6). As predicted by early models of U/V-SDR evolution[10], all seven genes are probably related to sex determination processes. Gametologue pairs include putative transmembrane proteins that may play a role in gamete recognition[30], STE20 serine/threonine kinase gametologues probably involved in pheromone pathways[31], and a casein kinase, a MEMO-like domain protein and a GTPase-activating protein which may act in signal transduction (Supplementary Table 6). All of these genes are gametologue pairs in *Ectocarpus* sp. 7, but in *D. herbacea* the casein kinase was lost from the U-SDR and the putative transmembrane receptor was lost in both sexes. We noticed that these ancestral V-SDR genes remain in the U/V-homologue of the species that have lost their U/V system (Fig. 2f and Supplementary Table 6), emphasizing their importance for pathways in sex organ development even in the absence of sex chromosomes.

The V-SDR size appears to be associated with the level of sexual dimorphism (Fig. 2e and Extended Data Fig. 2), but the small sample size is insufficient for formal statistical analysis. Species with low sexual dimorphism (anisogamous) retained the ancestral V-SDR genes with very few gene gains, further suggesting that they may represent the V-SDR ancestral state. Although the number of SDR changes is small, oogamous species each independently gained diverse V-SDR genes, and one gene (ATP-dependent RNA helicase) was convergently acquired in all (OG0003211 in Extended Data Fig. 2d). All the detected autosomal translocations into the V-SDRs of *Ectocarpus* sp. 7 and *D. herbacea* (Fig. 2e) also involve sex-specific genes (Supplementary Tables 2 and 5), consistent with a model where sexual antagonism in autosomal loci may be solved by gaining sex linkage[32]. In contrast, we found no correlation between autosomal sex-biased gene (SBG) expression and sexual dimorphism level (false discovery rate (FDR)-corrected $P > 0.01$; Supplementary Table 7), supporting previous studies[33] (Extended Data Fig. 5a). However, we observed an enrichment of male-biased genes on the PARs in all species (chi-square test $P < 0.01$) except *D. dichotoma* (Extended Data Fig. 5b).

Most U/V-SDR genes were prominently expressed in fertile haploid gametophytes, consistent with gene preservation via haploid purifying selection (Supplementary Table 8). Gametologues had typically higher expression levels than sex-specific genes (present in only one of the SDRs) (Wilcoxon test, $P = 0.00075$ in *D. herbacea*; $P = 0.08843$ in *Ectocarpus* sp. 7) (Fig. 2d). A comparative analysis in fertile gametophytes between SDR genes and their autosomal counterparts in other species showed that newly acquired genes on the SDR had similar expression levels to their autosomal counterparts (Extended Data Fig. 6), suggesting either a co-option of autosomal biological activity into male-specific functions in the V-SDR or the general importance of these genes for gametophyte development. Examining expression levels across multiple tissues in *Ectocarpus* sp. 7 revealed that activity of U/V-SDR genes is not confined to fertile gametophytes (Fig. 2d). Therefore, the SDRs contain not only genes involved in sex determination and gametophyte fertility but also genes playing a broader role in development.

Altogether, our analyses illustrate how brown algal U/V-SDRs undergo structural changes, evolving mainly by lineage-specific gene gains associated with increasing levels of sexual dimorphism. We identified a set of conservatively sex-linked genes in dioicous brown algae, suggesting their role in sex determination and/or differentiation, along with genes potentially involved in other developmental pathways.

## Structural features and evolutionary dynamics of brown algal U/V sex chromosomes

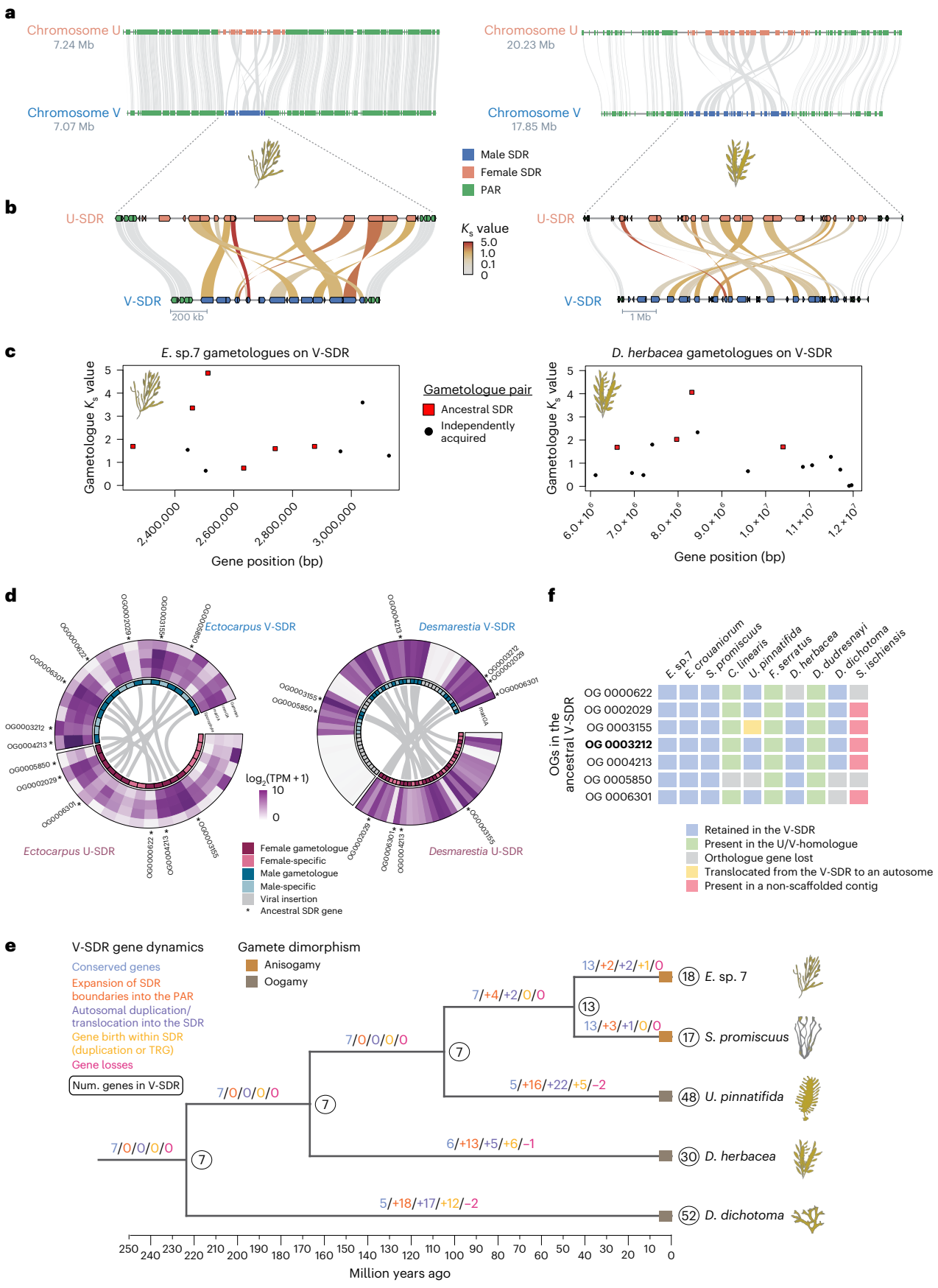
We next examined the structural features that differentiate the entire U/V sex chromosomes (V-SDR and PARs) from the rest of the genome (Fig. 3a and Supplementary Fig. 12a). As expected for non-recombining regions[25,34], all V sex chromosomes are repeat rich and gene poor (Wilcoxon rank-sum test, FDR-corrected $P < 0.01$; Fig. 3a, Extended Data Fig. 7 and Supplementary Tables 9–11). V-SDRs have significantly higher repeat density than the PARs or the autosomes (permutation test, FDR-corrected $P < 0.001$; Extended Data Fig. 8). This low gene density is not influenced by the presence of centromeres within the SDRs, as the coding density in the *Ectocarpus* sp. 7 V centromere (3.51%) is slightly higher than in the rest of the V-SDR (2.85%), presumably due to the small size of the centromere (153 kbp)[19]. The PARs were also significantly enriched in repeats when compared with the autosomes, although less so than the V-SDRs (permutation test, FDR-corrected $P < 0.001$; Extended Data Fig. 8). Among repetitive elements, 'unclassified' transposable elements (TEs) were enriched in the PARs and SDRs of the Ectocarpales (permutation test, FDR-corrected $P < 0.01$), while the V-SDRs of species that underwent genome expansion (for example, *U. pinnatifida*, *D. herbacea*, *D. dichotoma*) predominantly accumulated long terminal repeat (LTR) elements (Supplementary Fig. 6 and Supplementary Table 9).

Moreover, sex chromosomes had fewer orthologues conserved between species compared with the autosomes (chi-square test, $P < 10^{-4}$, Supplementary Table 12), possibly reflecting increased numbers of taxonomically restricted genes (TRGs; that is, genes that are not detectable outside of a defined taxonomic group). Phylostratigraphy analyses[35,36] confirmed an enrichment of TRGs in the sex chromosomes of all dioicous species (Wilcoxon rank-sum test, FDR-corrected $P < 0.01$; Fig. 3b and Supplementary Tables 13 and 14). TRG enrichment was localized in the PARs of the Ectocarpales, but this pattern extended to the entire sex chromosome, including the SDRs, in species with larger V-SDRs (permutation test, FDR-corrected $P < 0.001$; Fig. 3b). Importantly, sex chromosomes have statistically younger TRGs than the last common ancestor of the five dioicous species (same Order or broader taxonomic groups), indicating that TRG enrichment arose independently in each species (Pearson standardized residuals >2.4; Supplementary Figs. 7–11).

We previously proposed a theoretical model where generation-antagonistic selection may favour the retention of young sporophyte-beneficial loci in the PARs of *Ectocarpus* sp. 7 U/Vs[37]. Consistent with

**Fig. 2 | Lineage-specific U/V-SDR expansion from an ancestral SDR and its association with sexual dimorphism. a**, Microsynteny plot between the U and V chromosomes of *Ectocarpus* sp. 7 and *D. herbacea*. **b**, Synteny between the U and V gametologues within the V-SDRs of both species, coloured by synonymous substitutions per site ($K_s$). **c**, Identification of ancestral SDR gametologues (red squares) and independently acquired gametologues (black dots) with respect to their gametologue $K_s$ values and their position in the V-SDR. **d**, Circos plot linking gametologue pairs in each species with expression levels of all SDR genes ($\log_2$(TPM + 1)) across different life stages in *Ectocarpus* sp.7 and mature gametophytes (matGA) in *D. herbacea*. Gametologues are highlighted in dark colours, sex-specific genes are highlighted in light colours, viral insertions are marked in grey and asterisks denote conserved SDR genes (also in **f**). **e**, The ancestral state reconstruction of V-SDR gene content across brown algae, showing the expected number of genes in the SDR (white circles), gene retention (blue numbers), gene gain through expansion of the SDR boundaries (orange), gene gain through autosomal translocation (purple), gene birth event inside the SDR (yellow) and gene loss (red) along with changes in gamete dimorphism[16]. **f**, Schematic of the seven ancestral V-SDR gene orthogroups (OGs), with genomic locations marked: retained in the V-SDR (blue), found in the U/V-homologue of non-dioicous species (green), translocated from the V-SDR to an autosome (yellow), present in a non-scaffolded contig (red), and lost (grey). Bold: *MIN*. See Supplementary Tables 5 and 6.

https://doi.org/10.1038/s41559-025-02838-w



https://doi.org/10.1038/s41559-025-02838-w

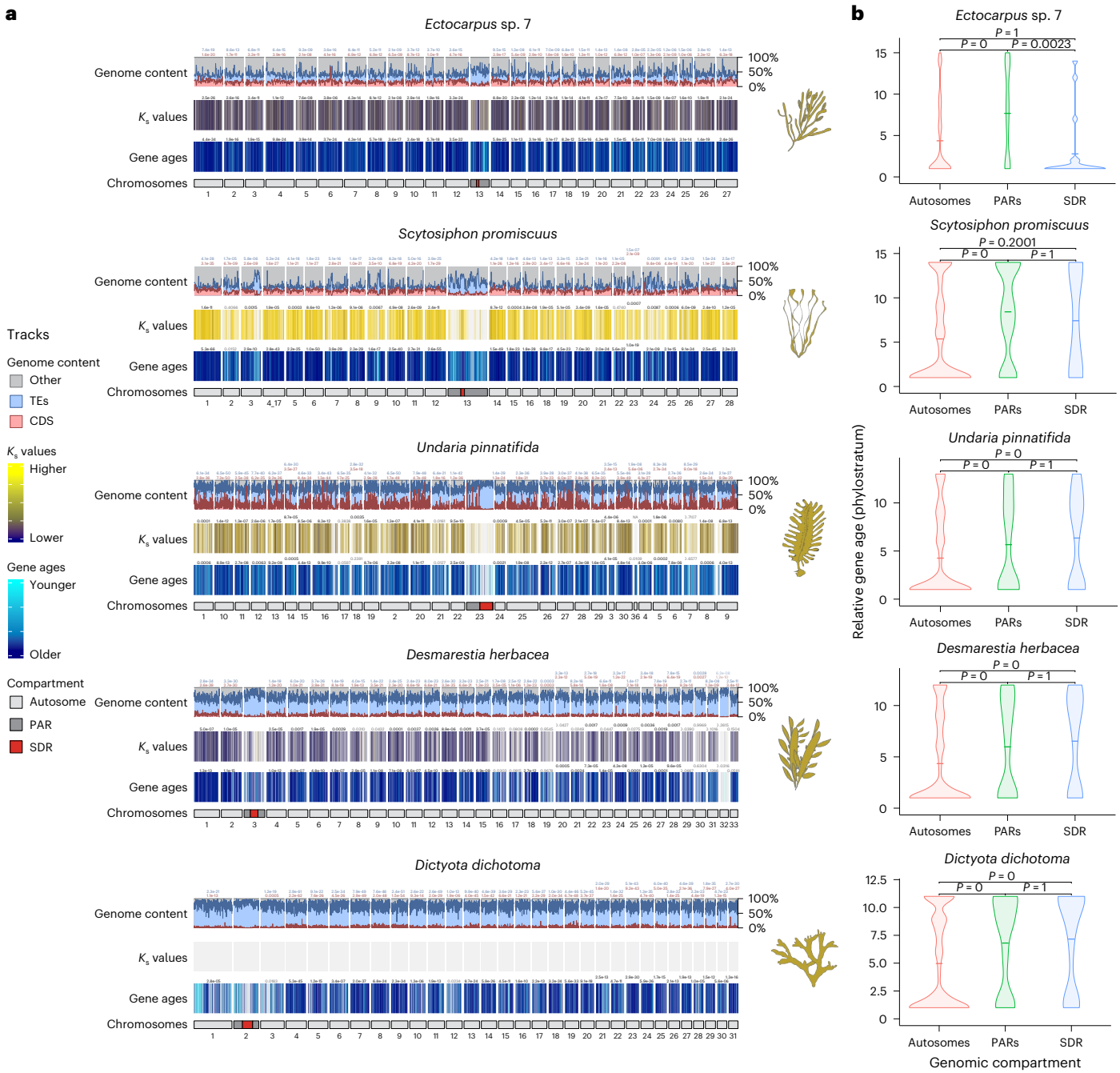



**Fig. 3 | The U/V sex chromosomes are enriched in taxonomically restricted genes. a**, Karyoplots for five dioicous species showing the following features from bottom to top: chromosome compartments (autosomes, PARs and SDR), relative gene ages, interspecies $K_s$ values, and proportion of coding (CDS, red) and repeat (TEs, blue) density. Statistically significant differences for each feature between each autosome and the V chromosome are depicted on top of the track for that autosome (FDR-corrected two-sided Wilcoxon rank-sum test; values indicated with solid colours when $P < 0.01$ for the tested hypothesis). The precise range of gene age categories and interspecies $K_s$ values for each species can be found in Supplementary Figs. 7–11. **b**, Violin plots for five dioicous species showing the relative gene age ranks (higher ranks equate to younger ages) of the TRGs across chromosome compartments (autosomes, PARs and SDR). Statistically significant differences in mean values of gene ages (centre line) were assessed using FDR-corrected two-sided permutation tests.

this model, sporophyte-biased genes are indeed enriched in the sex chromosomes of *Ectocarpus* sp. 7 and *U. pinnatifida* (fold change >2, adjusted $P < 0.05$), but less so in *D. dichotoma* and *S. promiscuus* (Supplementary Table 15). Moreover, we explored additional mechanisms underlying TRG emergence by estimating interspecies $K_s$ values between orthologues in closely related species (Supplementary Table 16) and comparing these values between chromosomes and genomic compartments (V-SDR, PAR, autosomes). If synonymous mutations behave neutrally[38,39], then interspecies $K_s$ can be used as a proxy for mutation rates[40,41]. Consistently, we found higher interspecies $K_s$ values in the V sex chromosomes compared with autosomes across all dioicous species (Wilcoxon rank-sum test, FDR-corrected $P < 0.01$; Fig. 3a and Supplementary Tables 16 and 17), suggesting higher mutation rates relative to autosomes. Higher interspecies $K_s$ values are also localized in the PARs, mirroring the pattern observed with the TRGs (Supplementary Figs. 7–10). Therefore, the enrichment of TRGs in the U and V is associated with both enrichment of sporophyte-biased genes and higher synonymous substitution rates.

To test the generality of the pattern of TRG enrichment on U and V chromosomes, we applied the same approach in other organisms with haploid sex determination, the plants *Ceratodon purpureum*, *Sphagnum angustifolium*, *Marchantia polymorpha*[11,12,42] and the fungus *Cryptococcus neoformans*[43]. We observed a clear enrichment of TRGs in the V chromosomes of *C. purpureum* and *S. angustifolium* (Wilcoxon rank-sum test, FDR-corrected $P < 0.01$; Supplementary Figs. 12 and 13 and Supplementary Tables 13 and 14), but not in the U/V chromosomes of *M. polymorpha* or the mating-type chromosome of *C. neoformans* (Supplementary Figs. 14 and 15 and Supplementary Tables 13 and 14).

### Fate of U/V sex chromosomes following loss of dioicy

We studied the evolutionary trajectory of brown algal genomes after the loss of the U/V system, by exploring two independent transitions to monoicy in *C. linearis* and *D. dudresnayi* that undergo sexual reproduction and develop male and female gametangia[33]. Most genes in the 'ex'-sex chromosomes (U/V-homologues) of both monoicous species are male derived, indicating that monoicy emerged from a male background (Fig. 4a,b). The U/V-homologue of *C. linearis* contains several rearrangements spanning the regions that are homologous to the PAR and SDR (SDR-homologue), with 11 V-SDR-derived and 2 U-SDR-derived orthologues located within the SDR-homologue, with an additional V-SDR-derived gene that was translocated elsewhere in the U/V-homologue (Fig. 4a and Supplementary Table 18). Likewise, *D. dudresnayi* underwent at least two inversion events within the SDR-homologue after splitting from *D. herbacea* (Fig. 4b), containing 20 V-SDR-derived genes and 4 U-SDR-derived genes (Supplementary Table 19).

Both monoicous species retained mostly male and a few female copies for most of the U/V-SDR-derived gametologues (91% in *C. linearis* and 100% in *D. dudresnayi*), whereas several U- and V-specific orthologues were lost in these species (10 and 17 sex-specific genes in *C. linearis* and *D. dudresnayi*, respectively). Of these lost orthologues, 60% and 43% present closely related autosomal paralogues in *C. linearis* and *D. dudresnayi*, respectively (Supplementary Tables 18 and 19), although it is unclear whether the expression of these autosomal paralogues is compensating the activity of the lost genes. The only three U-SDR-derived orthologues in *C. linearis* are flanked by PAR orthologues translocated at the end of the V-SDR-derived region, suggesting that the U/V-homologue (contig 12) of *C. linearis* acquired its U-SDR-derived genes through two translocations (Extended Data Fig. 9). Three U-SDR-derived genes in *D. dudresnayi* are dispersed across the V-SDR-derived region of the U/V-homologue, suggesting independent U-SDR translocations into the V-SDR, while the fourth U-SDR-derived gene was translocated to an autosome (Extended Data Fig. 9).

The seven ancestral V-SDR genes are transcriptionally active ($\log_2(\text{TPM} + 1) > 2$) during reproductive stages of both monoicous species (Supplementary Table 20), emphasizing their role in reproduction despite the absence of a U/V system, particularly *MIN*[22] which is retained in both species. While most U-SDR-derived genes are absent in monoicous species, a single intracellular cholesterol transporter gene was convergently preserved in both monoicous genomes (Supplementary Table 20) and actively expressed during fertility in both *Ectocarpus* sp. 7 and *D. herbacea* (Supplementary Table 8).

The U/V-homologue of *D. dudresnayi* retains some vestiges of its past as a U/V chromosome, such as low coding density, high repeat density and enrichment of TRGs (Wilcoxon rank-sum test, FDR-corrected $P < 0.01$), although we found non-significant differences in interspecies $K_s$ values across the genome (Fig. 4c, Extended Data Fig. 7, Supplementary Fig. 16 and Supplementary Tables 9–14, 16 and 17).

Finally, we examined the transition from haploid to diploid sex determination, which has remained unstudied in eukaryotes. Although the ancestral state for the brown algae is a U/V sexual system, the Fucales recently transitioned to a diploid life cycle[44], with many species, such as *F. serratus*[45], exhibiting diploid separate sexes (dioecy)[46]. Dioecy probably evolved from monoecy (both sexes in the same diploid individual) in the last common ancestor of the *Fucus* genus[47] (25–5 Ma[18]), consistent with a young sex chromosome in *F. serratus*. Our extensive bioinformatic analysis and PCR sex-linkage testing for candidate genes such as *MIN* (Methods) failed to identify sex-linked sequences in *F. serratus* (Supplementary Fig. 17), suggesting that the SDR is small and undifferentiated. However, male *F. serratus* conserves the *MIN* gene and all the ancestral V-SDR genes in its U/V-homologue (Figs. 2f and 5a). Importantly, although none of the U/V-SDR-derived genes are sex-linked in *F. serratus*, *MIN* and four other ancestral V-SDR genes are exclusively expressed in males (fully silenced in females) (fold change (FC) > 2, $p_{adj} < 0.05$; Fig. 5a and Supplementary Table 21). This pattern is consistent in three other Fucales species (Supplementary Table 22). Therefore, the ancestral V-SDR genes probably still play roles in male sex determination or differentiation pathways.

Contrary to the observations in *D. dudresnayi*, the U/V-homologue of *F. serratus* lacks the TRG enrichment pattern and all the other distinctive features of the U/V chromosomes (Wilcoxon rank-sum test, FDR-corrected $P > 0.01$). Thus, this 'ex'-sex chromosome has lost all the evolutionary vestiges of its past as a U/V chromosome (Fig. 5b, Extended Data Fig. 7, Supplementary Fig. 18 and Supplementary Tables 9–14, 16 and 17).

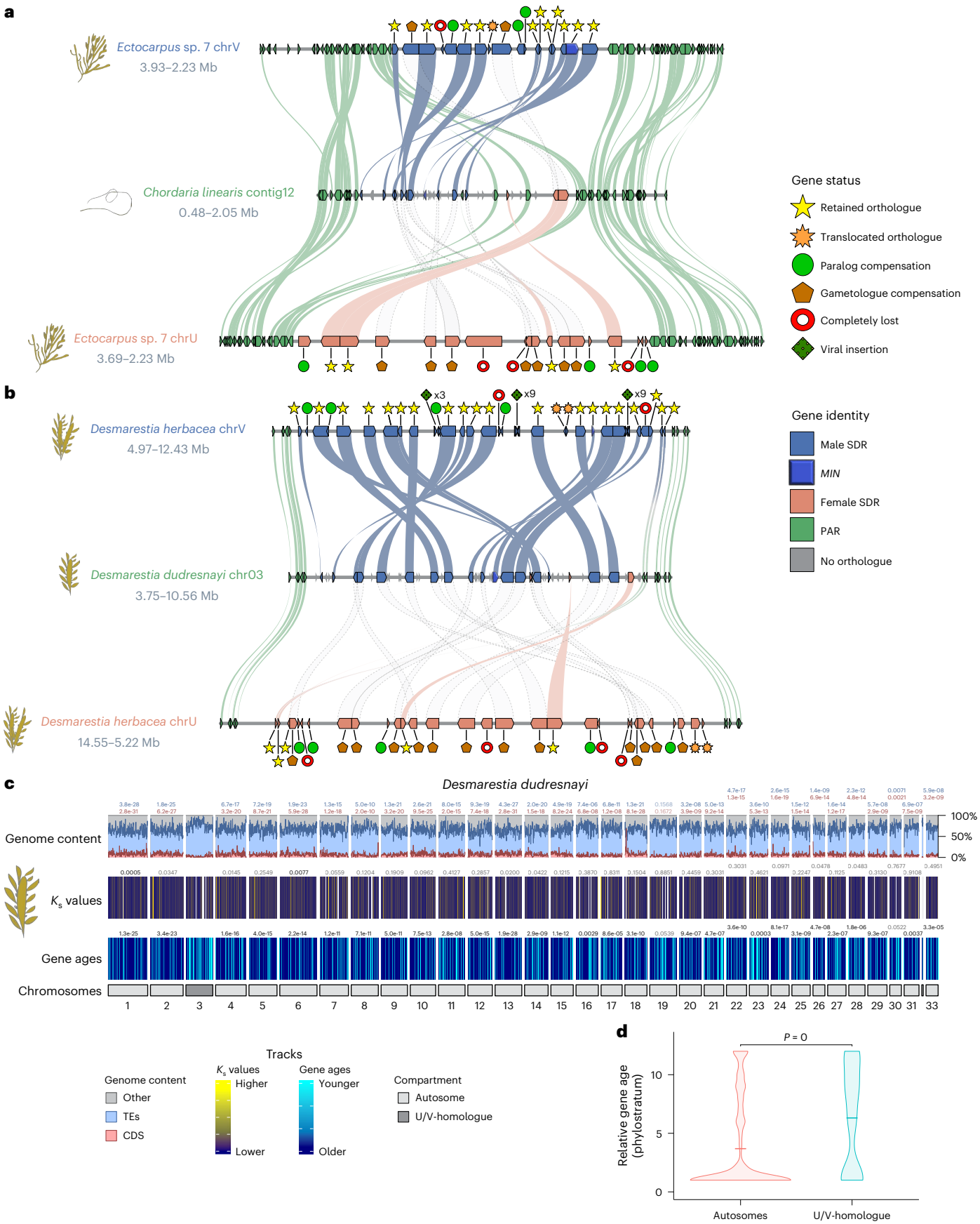
## Discussion

Here we characterized the evolutionary trajectory of brown algal sex chromosomes (Fig. 6). Brown algal sex chromosomes date back 450–244 Ma[21], at the origin of brown algae. We propose that the male-determining gene *MIN* underlies the birth of this U/V system[22]. The ancestral cassette with seven V-SDR genes suggests a very early evolution of these genes into a non-recombining locus during the evolution of the U/V-SDRs. The ancestral V-SDR genes probably contribute to reproduction, but may also be involved in broader developmental functions. Despite their old age, brown algal U/V chromosomes retain large PARs bordering the SDR, unlike haploid systems in non-vascular plants that mostly lack detectable PARs[11–14].

Brown algal genomes have a high degree of synteny conservation. The U/V-SDR is however prone to accumulate structural rearrangements, including inversions that may have caused the initial recombination suppression event in proto-sex chromosomes and the later expansion of the U/V-SDRs into the PARs. Similar to other haploid systems[48,49], TEs conspicuously accumulated in the SDRs following recombination suppression[50], possibly causing further rearrangements through TE-mediated inversions[51].

**Fig. 4 | Fate of sex chromosomes during transitions from dioicy to co-sexuality (monoicy). a**, Comparison of the U/V-homologue in *C. linearis* against the U and V chromosomes of *Ectocarpus* sp. 7. **b**, Comparison of the U/V-homologue in *D. dudresnayi* against the U and V chromosomes of *D. herbacea*. The colour code represents the identity of the genes alongside the chromosomes, while the shapes represent the evolutionary fate of each SDR gene in the monoicous genome. The matching shades between the SDRs and the U/V-homologue are either colour coded by their ancestral background or they appear as transparent dotted shades if the gametologue of the other sex was retained. **c**, Karyoplot of *D. dudresnayi* showing the following features from bottom to top: chromosome compartment (autosomes and U/V-homologue), relative gene ages, interspecies $K_s$ values, proportion of coding (CDS, red) and repeat (TEs, blue) density. Statistically significant differences for each feature between each autosome and the U/V-homologue are depicted on top of the track for that autosome (FDR-corrected two-sided Wilcoxon rank-sum test; values indicated with solid colours when $P < 0.01$ for the tested hypothesis). **d**, Violin plot showing the relative gene age ranks (higher ranks equate to younger ages) of the TRGs between the autosomes and the U/V-homologue of *D. dudresnayi*. Statistically significant difference in mean values of gene ages (centre line) was assessed using an FDR-corrected two-sided permutation test.

https://doi.org/10.1038/s41559-025-02838-w



https://doi.org/10.1038/s41559-025-02838-w

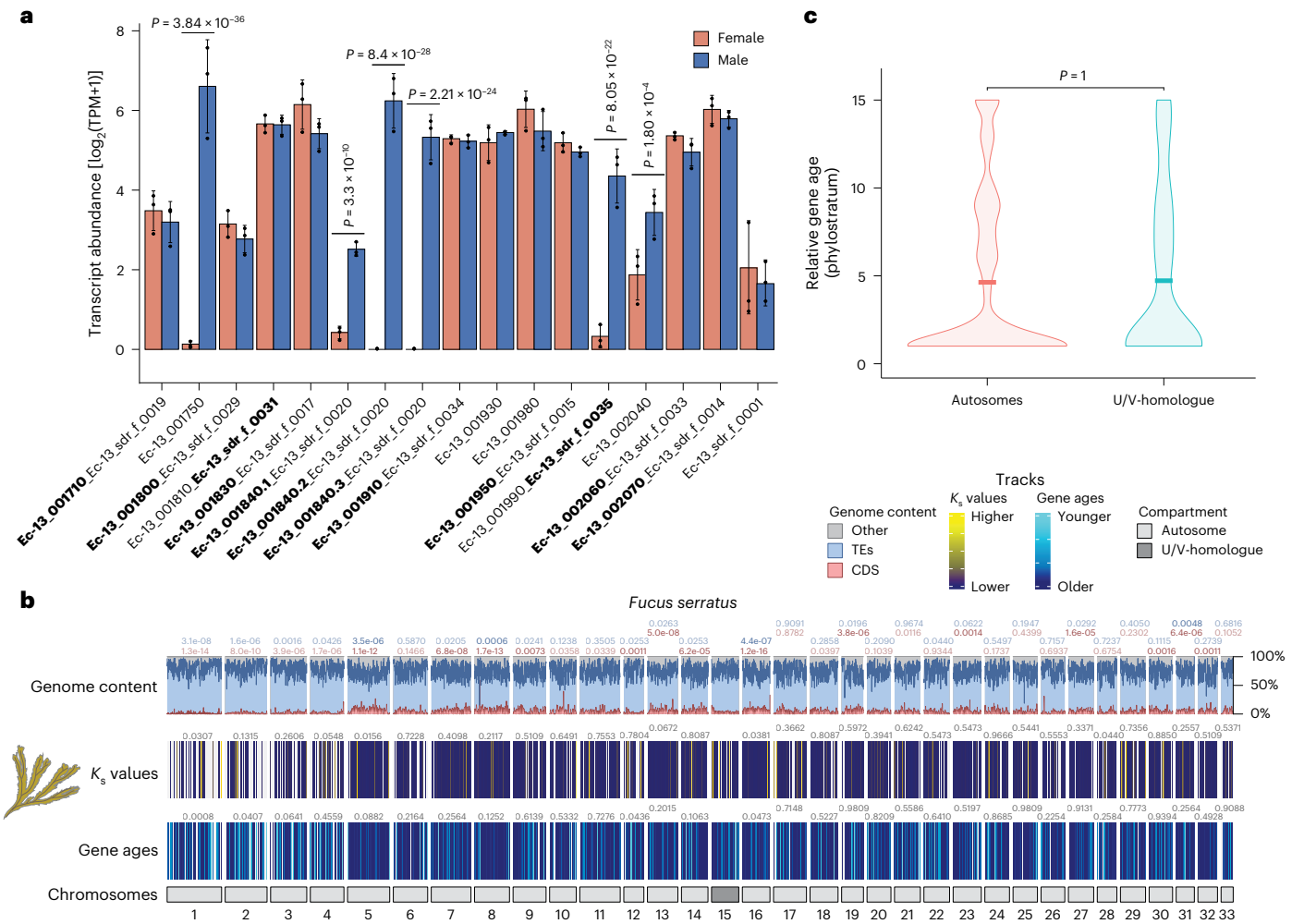



**Fig. 5 | Transition from haploid to diploid sex determination. a**, Expression of ancestral U/V-SDR genes in the diplontic species *F. serratus*. Gene expression of mature algae (using 3 males and 3 females, see Methods) is given as $\log_2(\text{TPM} + 1)$ and bars represent standard deviation of the mean. *P* values shown on the plot are derived from differential expression analysis performed using DESeq2, which applies a two-sided Wald test with Benjamini–Hochberg correction for multiple testing. Bold text represents whether the gene in *F. serratus* corresponds to an ancestral male or the female gametologue. **b**, Karyoplot of *F. serratus* showing the following features from bottom to top: chromosome compartment (autosomes and U/V-homologue), relative gene ages, interspecies $K_s$ values (between 0.00079 and 6.838, with an average value of 0.148), proportion of coding (CDS, red) and repeat (TEs, blue) density. Statistically significant differences for each feature between each autosome and the U/V-homologue are depicted on top of the track for that autosome (FDR-corrected two-sided Wilcoxon rank-sum test; values indicated with solid colours when $P < 0.01$ for the tested hypothesis). **c**, Violin plot showing the relative gene age ranks (higher ranks equate to younger ages) of the TRGs between the autosomes and the U/V-homologue of *F. serratus*. The mean values of gene ages (centre line) are not significantly different (FDR-corrected two-sided permutation test).

Models of XX/XY and ZW/ZZ sex chromosome evolution suggest that sexually antagonistic selection may lead to SDR expansions, making former PAR genes fully sex linked and gaining new genes[52,53]. We show that the three brown algal species with greater sexual dimorphism in gametes (oogamy) indeed have enlarged U/V-SDR gene contents, although statistical testing is not possible given the small sample size. Oogamy is proposed to be ancestral in the brown algae[16,21], but this trait seems to be highly labile and our analyses suggest multiple independent transitions to oogamy from a less dimorphic ancestor with a small ancestral SDR, accompanied by SDR gene gain. Anisogamous species only experienced few gene gains in their SDRs, as predicted in ref. 10. Similar to *M. polymorpha*, the relatively simple gene content of the U/V-SDRs in the brown algae could regulate an autosomal effector gene network controlling sexual dimorphism, including differences in somatic development between male and female gametophytes[54]. Accordingly, we observed substantial sex-biased expression of autosomal genes between mature male and female gametophytes carrying sex organs. This contrasts with *C. purpureus*, where sex chromosomes carry thousands of genes and few sex-biased autosomal genes[11]. Consistent with other studies[33,55], we found no correlation between levels of sex-biased expression and sexual dimorphism in brown algae. Sexual dimorphism in this lineage may be controlled by a relatively small subset of genes, and most sex-biased genes may influence gametophyte physiology or vegetative development.

In diploid sexual systems, recombination suppression between sex chromosomes explains Y and W degeneration[56] (but see refs. 57–59). Differences in the evolution of U/V systems are predicted, since sex determination occurs in the haploid stage, in which deleterious mutations are more efficiently removed by selection than in diploid systems[29,60] and some of the predicted differences have been documented in bryophytes[29]. In brown algae, as in diploid systems, the SDRs lost recombination, leading to TE accumulation and consequent reduction in gene density. However, unlike the degeneration documented in some XX/XY and ZW/ZZ systems, our ancestral state reconstruction revealed more gene gains than losses or gene movements out of the V-SDR during evolution of brown algal lineages.

https://doi.org/10.1038/s41559-025-02838-w

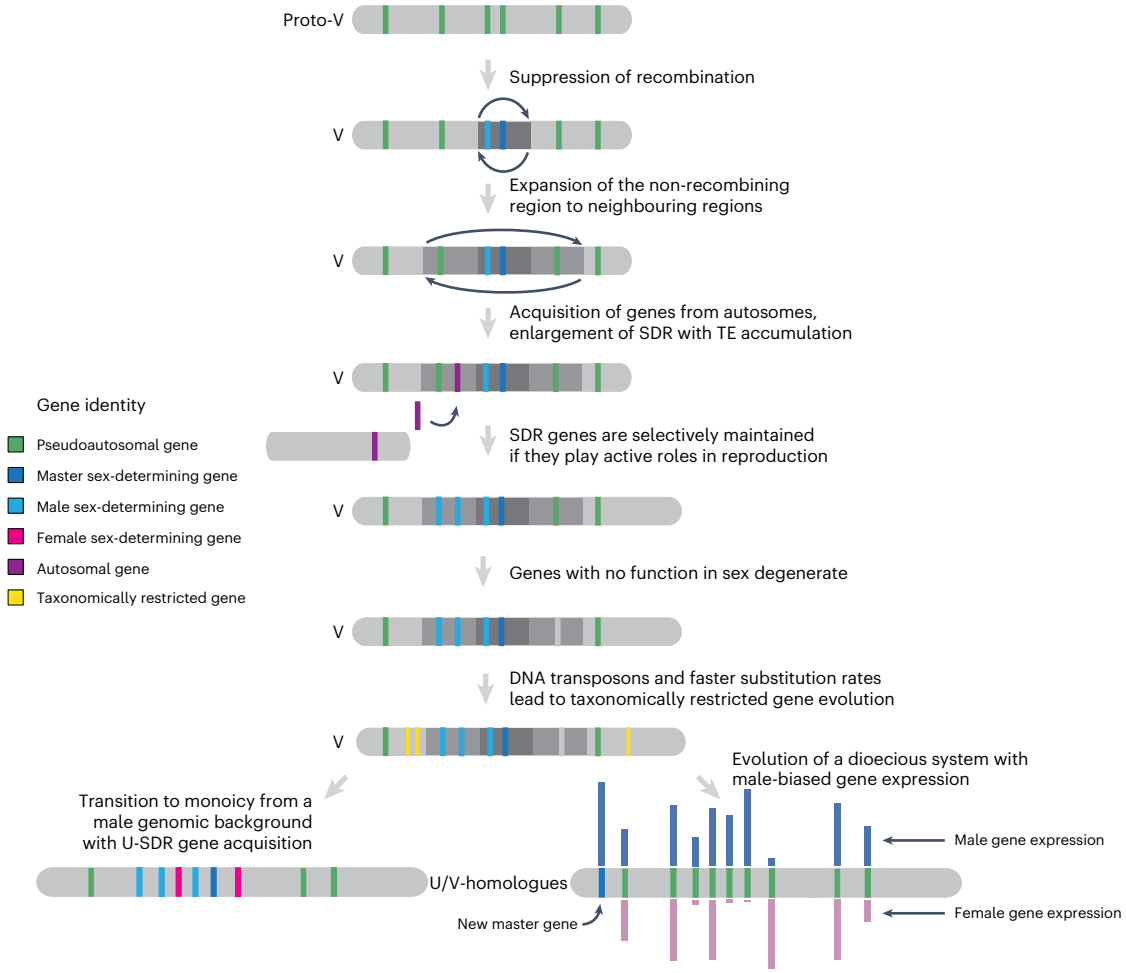



**Fig. 6 | Hypothetical model for U/V sex chromosome evolution.** U/V sex chromosomes arose from an ancestral autosome, via suppression of recombination that probably occurred via an inversion. The SDR boundaries expanded into neighbouring PAR via inversions, but also by recruitment of genes from autosomes; expansion occurred in a lineage-specific fashion, concomitant with increased sexual dimorphism of the different species. SDR genes are maintained within the SDR if they have roles in sex, whereas genes with no role in sex are lost. Faster substitution rates, probably a consequence of the heterochromatic context of the sex chromosome, may promote the rise of taxonomically restricted genes, which are selectively maintained on the sex chromosome if they have advantages to the sporophyte generation. In species that switch to a diploid life cycle, the U/V system disappears, but the genes that are in the V-specific region retain roles in sex, although they are no longer masters. Transition from U/V separate sexes to co-sexuality (monoicy) occurred when a male haploid individual acquired female-specific genes via translocations. During the demise of the U/V sex chromosomes, their structural and evolutionary footprints slowly disappear.

Unclassified repeats accumulate in the Ectocarpales sex chromosomes, including the PARs, but not in species with larger genomes, where LTR retroelements become dominant. DNA transposons are over-represented among unclassified repeats[61], and they often insert near the progenitor locus in a process called local hopping[62]. We propose that the U/V-SDR may thus act as a source of DNA transposons that hop to the PARs, thus increasing their repeat density, whereas increased colonization of LTR elements obscures this pattern in larger genomes.

Brown algal U/V chromosomes display an excess of TRGs. What mechanisms underlie this pattern? The SDR and the PARs of U/V chromosomes are enriched in heterochromatin[63], involved in repressing TEs[64], and heterochromatic regions tend to have higher mutation rates due to reduced access of the DNA repair machinery during replication[65]. Accordingly, we consistently observe higher interspecies $K_s$ values in U/Vs, particularly in the PARs, since these regions recombine between sexes and thus have a higher rate of neutral fixation than the SDR, which experiences the Hill–Robertson effect that reduces fixation probabilities of neutral mutations due to its linkage with other sites under selection[66]. We speculate that this feature could facilitate the evolution of TRGs. Alternatively, the high density of DNA transposons within the U/V could also promote the co-option of their regulatory motifs and enable de novo transcript birth, as seen in *Drosophila*[67]. Note that the pattern could be reinforced through generation-antagonistic selection[37], but DNA transposons and higher mutation rates may be sufficient to initiate this pattern in species lacking sporophyte-biased gene expression. Importantly, the TRG enrichment is unique to U/V systems and gradually disappears when these systems are lost. This pattern extends beyond brown algae to other eukaryotes with U/V systems, such as *C. purpureus*[11,68] and *S. angustifolium*[12]. However, we could not detect enrichment of TRGs on the small U/V sex chromosomes of *M. polymorpha*[42,69], or on the mating-type chromosome of *C. neoformans*[43]. Mosses such as *C. purpureus* and *S. angustifolium* display relatively more complex sporophyte body plans than liverworts such as *M. polymorpha*[70], which could underlie stronger generation-antagonistic selection[37]. Unlike brown algae, most of the sex chromosomes in bryophytes are sex linked, with minimal space for PARs[29]. In this context, the sex chromosome of *C. purpureus* expanded its SDR very recently in evolutionary time through fusions of autosomes with earlier established U and V chromosomes[71], retaining its evolutionary footprints as PARs,

while the sex-linked region in *M. polymorpha* is much older[69], which could also limit the formation of TRGs that are predominantly observed in the PARs of the brown algae. The pattern thus appears to be specific to U/V systems where chromosomal degeneration is mild and linked to haploid–diploid life cycles where the sporophyte stage is sufficiently complex, highlighting a key role for generation selection[37]. Therefore, our study hints at a unique interplay between complex life cycles, heterochromatic landscape, DNA transposons and higher mutation rates that may lead to TRG enrichment in U/V chromosomes, and this process is pervasive across distant, independently evolved eukaryotic kingdoms.

Monoicous brown algae have transcriptomic profiles resembling ancestral females[33]. However, our results show that monoicy arose at least twice from a male ancestor that acquired female genes. The male pathway requires *MIN*[22,25,72], which could have facilitated the evolution of monoicy from males, as also seen in the green lineage[73,74]. We note the presence of U-SDR-derived gene(s) in all monoicous species, particularly a cholesterol transporter gene that is found in all U-SDRs, suggesting that the U-SDR contains femaleness-promoting factor(s), consistent with reports in kelps[75]. We thus propose that monoicy evolved via translocation events adding essential femaleness-promoting genes to a male genetic background. Since the resulting monoicous individuals were capable of producing both male and female reproductive structures, individuals with U chromosomes were no longer essential for sexual reproduction, ultimately leading to the loss of the U-SDR in monoicous species. Although the combination of key female and male genes is essential for this evolutionary transition, the retention of a sex chromosome is not. For example, in *Volvox africanus*, monoicy required the retention of female SDR-like regions, while most male SDR genes were lost except for a multicopy array of the male-determining gene *MID*[76].

The evolution of a dioecious system in *F. serratus* is associated with an irreversible transition to diploidy in Fucales[16,47]. The U/V to dioecy transition has remained elusive[2,47], but our data in brown algae imply that it involved an intermediate monoicous stage, supporting previous predictions from ancestral state reconstruction analyses[16]. A small, undifferentiated Y-specific region consistent with a young XY system may explain why the sex chromosome in *F. serratus* was undetectable. Nonetheless, all ancestral V-SDR genes are found in the U/V-homologue of *F. serratus*, several showing a male-biased gene expression across Fucales species, particularly *MIN*[22]. Our findings imply that *MIN* and possibly other ancestral V-SDR genes are still involved in male differentiation, but shifted downwards in the sex determination cascade. These results thus support and extend the 'bottom-up' hypothesis of sex determination, where downstream components of sex differentiation are conserved across taxa, and new master sex regulators can replace older ones[77].

## Methods

### Biological material

*Scytosiphon promiscuus, Dictyota dichotoma*, *Undaria pinnatifida* and *Desmarestia dudresnayi* haploid gametophytes were cultivated under laboratory conditions as in ref. 78. We cultivated the gametophytes at 14 °C with a photoperiod of 12:12 h light:dark at an irradiance of 25 µmol photons $m^{-2}$ $s^{-1}$. The media consisted of filtered natural seawater (NSW), which was autoclaved and enriched with half-strength Provasoli nutrient solution (Provasoli-enriched seawater; PES)[78]. We grew the first biomass in 140 mm Petri dishes and the gametophytes were later transferred to a 1 l flask with gentle aeration. The gametophytes were fragmented once a month and the media were changed every 2 weeks to promote biomass production. Before freezing, gametophytes were treated with antibiotics for 3 days with gentle agitation and under the same culture conditions. The first day, gametophytes were treated with a mix of streptomycin (2 g $l^{-1}$ of PES), penicillin G (0.5 g $l^{-1}$ of PES) and chloramphenicol (0.1 g $l^{-1}$ of PES); the next day with ampicilin (1 g $l^{-1}$ of PES), and on the last day with kanamycin (1 g $l^{-1}$ of PES). Between each day of treatment and before freezing, gametophytes were rinsed with 500 ml of NSW to remove traces of antibiotic.

Samples of fucoid algae sexual and vegetative tissue were collected in the intertidal zone during low tides in June 2012 from Viana do Castelo (*F. vesiculosus*, *A. nodosum*) and Caminha (Rio Minho; *F. ceranoides*), northern Portugal. Sexual phenotypes were verified in the field by sectioning and observing receptacles under a field microscope. Tissue samples were flash frozen in liquid nitrogen on the shore and transported to the laboratory in a cryoshipper, after which they were lyophilised and stored dry at room temperature on silica crystals (see Supplementary Table 1 for a list of strains used in this study).

### DNA and RNA extraction and sequencing

Genomic DNA was isolated from algal tissue (~100 mg) by grinding into fine powder under liquid nitrogen and subsequent cell lysis in 500 µl of Genomic Lysis Buffer (OMNIPREP for plant kit) for 1 h at 60 °C. The lysate was cleaned up with 200 µl of chloroform and DNA was precipitated in ethanol. The DNA pellet was digested in CF buffer (Macherey–Nagel) for 45 min at 65 °C and purified using NucleoBond AXG20 Mini columns according to the user manual (Macherey–Nagel). Final high molecular weight genomic DNA was quantified (Qubit), analysed for purity (Nanodrop) and checked for size distribution (Femto Pulse System) before preparing the sequencing libraries. We sequenced the libraries using an Oxford Nanopore Technologies (ONT) MinION Mk1B. We prepared the ONT libraries using an SQK-LSK110 library preparation kit for R9.4.1 flow cells and an SQK-LSK114 library preparation kit for R10.4.1 flow cells. Two libraries were sequenced for *D. dudresnayi* on R9.4.1 flow cells and a third library was sequenced on an R10.4.1 flow cell.

RNA was isolated from mature gametophytes of *U. pinnatifida* and *S. promiscuus* following modified procedure of the Qiagen RNAeasy kit, and the TruSeq RNA Library Prep Kit v.2 was used to sequence the transcriptomes in an Illumina NextSeq 2000 platform (150 bp, PE reads). Extraction of total RNA from fucoid algae (*F. vesiculosus*, *A. nodosum* and *F. ceranoides*) was performed following ref. 79 and RNA libraries were sequenced on an Illumina HiSeq 2000 machine (100 bp, PE reads).

### Genome assembly and annotation

High-quality, chromosome-level assemblies of brown algal genomes have been notoriously difficult to obtain due to technical challenges in extracting nucleic acids. Whole-genome assemblies and annotations of *S. promiscuus* male, *D. dichotoma* male, *D. herbacea* male and female, *E. crouanorium* male, *C. linearis*, *S. ischiensis* and *F. serratus* male were obtained from ref. 18. We also downloaded the genome of *Ectocarpus* sp. 7 (ref. 19) and the male genome of *U. pinnatifida*[20], which were already assembled at a chromosome level. For *D. dudresnayi*, we performed genome sequencing, de novo genome assembly and ab initio gene annotation. Base calling was done using ONT Guppy[80] with the configuration files 'dna_r9.4.1_450bps_sup.cfg' and 'dna_r10.4.1_e8.2_400bps_sup.cfg' and the options '–trim_adapters –trim_primers', yielding 17.4 Gbp of data in 2,871,152 reads. We merged all the reads and analysed them using Kraken (v.2.1.2)[81] and the bacteria database (August 2022) to remove potential contaminant sequences. All data classified as bacterial reads by Kraken were screened using blastN (v.2.13.0+)[82] (-evalue 0.001 -num_alignments 20) against the NCBI genbank bacterial database (downloaded November 2023). The blastN output was visualized in MEGAN (v.6.23.4)[83], and all reads that were declared as bacterial were extracted and removed from further analyses. We obtained 1,908,772 decontaminated reads with an average length of 5.1 Kbp (9.8 Gbp of data, 20× coverage), which were deposited on the NCBI Sequence Read Archive (Supplementary Table 1).

The decontaminated reads were assembled de novo using flye (v.2.9.1-b1780)[84] with the options '–nano-raw -g 450 m -t 28 -i 3 –scaffold'. The draft assembly consisted of 1,032 contigs with a total size of 425 Mbp, an N50 of 4.6 Mbp and an L50 of 29 contigs. We used TransposonPSI (http://transposonpsi.sourceforge.net/) to predict

the TEs and RepeatScout (v.1.0.6)[85] to predict the simple repeats in the genome assembly. Both predictions were combined to soft mask the repetitive content in the genome assembly using bedtools maskfasta (v.2.27.1)[86]. We mapped the RNA-seq data of *D. dudresnayi* from the PhaeoExplorer database[18] to the soft-masked genome assembly using STAR (v.2.7.1a)[87]. We used BRAKER v.2.1.6 alongside the RNA-seq data[88] to predict the protein-coding genes in the soft-masked genome assembly.

### Hi-C library preparation and sequencing for chromosome-level assemblies

We generated Hi-C libraries for three male genomes (*S. promiscuus*, *D. herbacea* and *D. dichotoma*) and two female genomes (*Ectocarpus* sp. 7 and *D. herbacea*). Fresh algal tissue was cross-linked for 20 min at room temperature in a solution of 2% formaldehyde with filtered NSW and then transferred into a 400 mM glycine solution with filtered NSW for 5 min to quench the formaldehyde. The samples were then stored at −80 °C until use. The Hi-C libraries were prepared as follows. The samples were de-frosted in 1 ml of 1× *Dpn*II buffer with protease inhibitors (Roche cOmplete), transferred to Precellys VK05 lysis tubes (Bertin) and disrupted using the Precellys apparatus with five grinding cycles of 30 s at 7,800 r.p.m., followed by 20 s pauses. SDS was added to the lysate at 0.5% final concentration and samples were incubated for 10 min at 62 °C, followed by the addition of Triton X-100 to a final concentration of 1% and 10 min of incubation at 37 °C under gentle shaking. We added 500 U of *Dpn*II to 4.6 ml of the digestion mixture and incubated the samples for 2 h at 37 °C under gentle shaking (180 r.p.m. in an inclined rack to prevent sedimentation), followed by the addition of another 500 U of *Dpn*II and an overnight incubation under the same conditions. The digested samples were centrifuged at 4 °C for 20 min at 16,000 × *g*. The supernatant was discarded and the pellet was incubated for biotinylation at 37 °C for 1 h under a constant shaking (300 r.p.m.) in a 500 ml biotinylation mix with a concentration of 1× ligation buffer, 0.09 mM dATP-dGTP-dTTP, 0.03 mM biotin-14-dCTP and 0.64 U $ml^{-1}$ Klenow fragments. After biotinylation, the samples were incubated for 3 h at room temperature in a 1.2 ml ligation reaction with a concentration of 1× ligation buffer, 100 mg $ml^{-1}$ BSA, 1 mM ATP and 0.4 U $ml^{-1}$ T4 DNA Ligase. The samples were then incubated overnight at 65 °C after adding 20 µl 0.5 M EDTA, 80 µl 10% SDS and 1.6 mg Proteinase K. DNA was extracted with 1 volume of phenol/cholorform/isoamyl alcohol (24:24:1), followed by 30 s of vortexing at top speed and a 5-min centrifugation at top speed. We precipitated the DNA by adding 1/10 volume of 3 M NaAC pH 5 and two volumes of 100% cold ethanol, followed by a 30-min incubation at −80 °C and a 20-min centrifugation at 14,000 × *g* and 4 °C. The DNA pellet was washed with 1 ml 70% ethanol, then dried at 37 °C for 10 min and resuspended in 100 µl 1× TE buffer with 1 mg $ml^{-1}$ RNase. DNA was sheared to 250–500 bp fragments using a Covaris S220 ultrasonicator, purified with AMPure beads (0.6×) (Beckman) and eluted in 20 µl 10 mM Tris pH 8.0. Biotinylated but not ligated DNA fragments were first removed by T4 DNA polymerase treatment (final concentration, 300 U per pellet; NEB), and the biotin-labelled fragments were selectively captured by Dynabeads MyOne Streptavidin C1 beads (Invitrogen). The libraries were prepared using the NEB Ultra II library preparation system and sequenced on the NextSeq 2000 Illumina platform (2 × 150 bp) (Supplementary Table 1).

We scaffolded the genomes from ref. 18 into chromosome-level assemblies using the Hi-C data. We filtered the low-quality Hi-C reads using Trimmomatic (v.0.39)[89] (ILLUMINACLIP:2:30:10 LEADING:25 TRAILING:25 SLIDINGWINDOW:4:15 MINLEN:75 AVGQUAL:28). We mapped the Hi-C reads against each genome assembly using BWA-mem v.0.7.17-r1188d in the Juicer v.1.6 pipeline[90] to generate a contact map, which was then fed to 3D-DNA v190716 (ref. 91) to scaffold the genomes into chromosomes. The obtained scaffolds were manually inspected against the contact maps to solve the limits of each chromosome using Juicebox (v.1.11.08)[92]. The PhaeoExplorer gene annotations[18] were lifted into the new assemblies using Liftoff (v.1.6.1)[93], while the annotation of TEs was performed using RepeatModeler2 (ref. 94). We scaffolded the genomes of *E. crouaniorum* and *D. dudresnayi* into chromosomes using a reference-guided assembly with RagTag (v.2.0.1)[95] against the chromosome-level assemblies of *Ectocarpus* sp. 7 and *D. herbacea*, respectively. All genes within the SDRs in the brown algal species studied (see below) were manually curated to exclude any TE-related genes from the annotation.

### Discovery of the U/V sex determination regions

Male sex-determining regions (V-SDR) in *S. promiscuus*, *U. pinnatifida*, *D. herbacea* and *D. dichotoma*, as well as female sex-determining region (U-SDR) in *D. herbacea* were analysed following two complementary methods: (1) a *k*-mer-based YGS approach, originally designed to detected Y-linked sequences in heterogametic systems, developed in ref. 96 and (2) genomic coverage analysis, designed to identify sex-linked regions through differences in read depth between male and female individuals[97]. These methods are well suited for organisms with divergent sex chromosomes, such as brown algae, where U and V haplotypes have diverged over extended evolutionary time.

The YGS method principle is to identify male or female sex-linked scaffolds by comparing *k*-mer frequencies between reference genome assembly and *k*-mers generated from DNA-seq reads of the opposite sex. Regions in the male reference genome that contain *k*-mers that are absent in female reads will indicate candidate male SDR sequences; similarly, female genomic scaffolds with low coverage in male *k*-mers will denote female SDR region. For each species, 15-base-pair *k*-mer sequences were generated separately from male and female Illumina reads (see Supplementary Table 1 for data accession numbers) using Jellyfish v.2.3.0 count (-m 15 -s 10 G -C –quality-start=33 –min-quality=20) and converted to fasta format with Jellyfish dump (–lower-count=5)[98]. Next, non-overlapping 500-kb sliding windows (*Desmarestia*, *Dictyota* and *Undaria*) or 200-kb sliding windows (*Scytosiphon*) of the reference chromosome genomes (from the sex whose SDR was to be identified) were created using seqkit (v.2.3.1)[99] and used as input for the YGS.pl script[96] together with the fasta *k*-mer files produced in the previous step. Each window was then analysed to calculate the proportion of *k*-mers in the reference window that are not present in the opposite-sex *k*-mer database. Genomic windows with a minimum of ≥50% of unmatched single-copy *k*-mers were then retained as candidate male or female SDR sequences. Because the borders of the SDRs cannot be precisely defined at the single-nucleotide level with the available data, we focused on genes within these regions and defined the SDR boundaries on the basis of the flanking genes located at the transition to pseudoautosomal regions (PARs).

Candidate SDR regions identified by YGS were further validated by analysing sex-specific differences in read coverage. In detail, the short Illumina reads coming from males and females of each investigated species were trimmed with Trimmomatic[89] (see above) and mapped to the reference genome for which the SDR was to be studied, using HISAT2 (ref. 100) (default settings). Bam files produced by HISAT2 were used as input for Mosdepth[101] to calculate coverage in 10-kb windows along the genome sequence (-m -n -b 10000 –fast-mode -Q 30). Read mapping depth in genomic windows was normalized by the genome-wide mean for each sex, and the coverage in genomic intervals was then compared between males and females. Because V-SDR-linked sequences are present only in males, we expect them to have similar read coverage as autosomal regions in males, but little or no coverage in females (and conversely for the U-SDR sequences in *D. herbacea*). The comparison focused on regions within male reference genomes where the coverage in males fell within the range of 75–125% of the genome average, while the coverage in females remained below 50% of the genome average.

Both coverage and *k*-mer analysis identified identical genomic regions, providing high-confidence candidate SDRs (Extended Data Table 1 and Supplementary Figs. 1–5). In *D. herbacea*, where both male and female chromosome-level genome assemblies were available,

we directly compared U and V chromosomes to further confirm the SDR borders by analysing the collinearity of pseudoautosomal regions flanking the SDRs. The SDR scaffolds for all studied species were further validated by PCR amplification (Supplementary Table 1) using 4 males and 4 females.

### Genetic mapping and search for the sex chromosome in *F. serratus*

Three different sets of materials were used in this study: (1) 12 male and 12 female field samples, hereafter denoted the 24-individual natural population; (2) 157 sporophyte progeny population derived from a cross between one male sample and one female sample collected from the field and (3) 3 male and 3 female samples collected from the field for whole-genome sequencing. The 157-progeny population and 24-individual natural population were genotyped using the double digest RAD sequencing approach (ddRAD-seq). Briefly, individual genomic DNA was digested with the restricted enzymes PstI and HhaI to obtain fragments that were size selected between 400 and 800 bp before sequencing on in Illumina HiSeq 2500 platform (paired-end 2 × 125 bp). See ref. 102 for detailed protocol of the ddRAD-seq.

We performed whole-genome sequencing on an Illumina HiSeq 2500 system (2 × 150 bp paired-end) for the 3 male and 3 female samples. For ddRAD-seq data, raw reads were cleaned and trimmed with Trimmomatic as above and mapped to the draft genome of *F. serratus* male. For the progeny population, genotypes were called from the obtained bam files using the Stacks pipeline (v.2.5)[103]. The obtained vcf files were filtered with VCFtools (v.0.1.16)[104] and bcftools[105] (max missing per locus: 30%, max missing per sample: 40%, max mean coverage:30, minQG:20).

The filtered vcf file of the progeny population was used to construct a genetic map with Lep-MAP3 (ref. 106). Briefly, the ParentCall2 module was used to call parental genotypes, the SeparateChromosomes2 module was used to split the markers into linkage groups and the OrderMarkers2 module was used to order the markers within each linkage group using 30 iterations per group and finally computing genetic distances. Phased data were converted to informative genotypes with the script map2genotypes.awk.

### We used different approaches to identify the SDR in *F. serratus*.

*Coverage analysis*. We combined whole-genome sequence data from the 3 males and 3 females alongside the ddRAD-seq data of the 24-individual natural population, mapping both datasets to the *F. serratus* male genome assembly using bwa-mem[107]. Coverage analyses was done in several ways:

- Using SATC (sex assignment through coverage)[108], a method that uses sequencing depth distribution across scaffolds to jointly identify: (1) male and female individuals and (2) sex-linked scaffolds. This identification was achieved by projecting the scaffold depths into a low-dimensional space using principal component analysis and subsequent Gaussian mixture clustering. Male and female whole-genome sequences were used for this analysis.
- Using the method SexChrCov described in ref. 109 with the 24-individual natural population.
- Using the method DifCover[110] which identifies regions in a reference genome for which the read coverage of one sample is significantly different from the read coverage of another sample when aligned to a common reference genome. The 24-individual natural population was used for this analysis.
- Using soap.coverage (v.2.7.9)[111] to calculate the coverage (number of times each site was sequenced divided by the total number of sequenced sites) of each scaffold in each sample. For each scaffold, the male to female (M:F) fold change coverage was calculated as $\log_2$(average male coverage) − $\log_2$(average female coverage). The 24-individual natural population was used for this analysis.

*Fixation index ($F_{ST}$) and sex-biased heterozygosity*. This approach has been previously used to find sex-linked genomic regions in several studies[112,113]. Using the 24-individual natural population, $F_{ST}$ was calculated using vcftools[104]. Sex-biased heterozygosity was defined as the $\log_{10}$ of the male heterozygosity:female heterozygosity ratio, where heterozygosity was measured as the fraction of sites that are heterozygous. This ratio is expected to be zero for autosomal scaffolds and elevated on young sex scaffolds due to excess heterozygosity in males.

*Identification of eventual female scaffolds that failed to map to the male reference genome*. Vcftools and bedtools were used to extract female regions that did not map to the reference genome consistently in the 3 resequenced female samples.

All candidate contigs were tested by PCR in 4 males and 4 females.

### Synteny analyses, $K_s$ analysis and transitions to co-sexuality

Whole-genome synteny comparisons were performed for each pair of chromosome-level assemblies using MCscan (v.1.2.14)[114], both between different species, between sex chromosomes in the same species and between monoicous species and their closest relatives with U/V chromosomes. The putative gametologues between sex chromosomes that were predicted with MCscan were reassessed using OrthoFinder (v.2.5.4)[115] and best reciprocal DIAMOND (v.2.1.8.162)[116] hits.

We calculated the number of synonymous substitutions per synonymous site ($K_s$) for each pair of male and female gametologues as a proxy to assess the relative time at which both genes diverged from each other. The amino acid sequences of each pair of gametologues were aligned with MAFFT (v.7.520)[117] and subsequently aligned into codons using pal2nal (v.14)[118]. The gametologue $K_s$ values were calculated using the model in ref. 119 as implemented in KaKs_calculator (v.2.0)[120].

We evaluated the male or female identity of the genes in the monoicous species whose orthologues were found within the SDR in their closest dioicous relatives. For this, we compared the results obtained with MCscan[114] against the orthogroup prediction performed with OrthoFinder[115], with best reciprocal DIAMOND[116] hits and by calculating gene trees for each orthogroup using an amino acid alignment with MAFFT[117] and gene tree reconstructions using FastTree (v.2.1.11)[121].

### Ancestral reconstruction of the male SDR

The brown algal phylogeny was obtained from ref. 18. The species tree is based on 32 single-copy nuclear genes whose protein sequences were aligned manually using AliView[122], and whose best-fit substitution models were assessed independently using the Akaike information criterion. The tree was generated using a maximum likelihood approach implemented in RAxML bootstraps and the gamma model. Every node in the phylogeny has 99–100% bootstrap support values. Divergence times were subsequently calculated using MCMCtree[123] and three calibration points. The MCMC chains were run for 1.5 million generations and the first 200,000 MCMC chains were discarded as burn-in.

We searched for orthologue genes within the V-SDR of five species (*Ectocarpus* sp. 7, *S. promiscuus*, *U. pinnatifida*, *D. herbacea* and *D. dichotoma*) in our OrthoFinder results. For each V-SDR gene, we coded its orthologue in the other species as 'present' (1) if it is also sex linked in the V-SDR, whereas it was coded as 'absent' (0) if the orthologue resides in the PARs, in an autosome or if there is no detectable orthologue in that species. Once we generated this presence/absence matrix with the evolutionary relationship of the genes within the V-SDR (Supplementary Table 5), we used it as the input file for the software Count (v.10.04)[124] to estimate the ancestral content of the V-SDR throughout a phylogeny and determine the most likely scenario of V-SDR evolution in the brown algae. We employed posterior probabilities under a phylogenetic birth-and-death model with independent gain and loss rates across each branch in the phylogeny. We modelled the independent gain and loss rates through 10 gamma categories and performed 1,000 optimization rounds with a convergence threshold

on the likelihood >0.1 to find the best fitting model for the data. The branch lengths in the tree that were used for the ancestral state reconstruction were retrieved from the molecular clock analysis performed in ref. 16. We distinguished between conserved V-SDR genes that are ancestral and parallel acquisitions of the same gene in the V-SDR by analysing gene trees between male and female genomes, in addition to female transcriptome assemblies of *D. dichotoma* and *U. pinnatifida*. Sequence alignments were done using MAFFT[117] with default settings and uploaded to the https://www.phylogeny.fr/ platform. Alignments were further curated using Gblocks (v.0.91b)[125] (min. seq. for flank pos.: 85%, max. contig. nonconserved pos.: 8, min. block length: 10). Trees were produced using PhyML (v.3.1l)[126] with the default model and visualized in TreeDyn (v.198.3)[127]. The approximate likelihood-ratio test was chosen as the statistical test for branch support. We inferred the function of the ancestral V-SDR genes through the annotation of genes in *Ectocarpus* sp. 7 belonging to that orthogroup. The most likely acquisition mechanism of each SDR gene in each species was assessed on the basis of the position of each orthologue in the other species (pseudoautosomic, autosomic or missing; Supplementary Table 5).

### Genomic content across chromosomes

We used closely related genome assemblies available in the PhaeoExplorer database[18] to assess the depletion of orthologues in the sex chromosome. We predicted one-to-one orthologues using OrthoFinder[115] between the following species pairs: *Ectocarpus* sp. 7 with *Ectocarpus siliculosus*, *S. promiscuus* with *C. linearis*, *U. pinnatifida* with *Saccharina japonica*, *F. serratus* with *Fucus distichus*, *D. herbacea* with *D. dudresnayi*, and *D. dichotoma* with *Halopteris paniculata* (Supplementary Table 16). We calculated the expected number of detectable orthologues for each chromosome and compared it against the observed number of detected orthologues using chi-square tests. We performed Benjamini–Hochberg corrections to the *P* values of the chi-square tests to control the FDR in the analysis[128].

GenEra[36] was used by running DIAMOND in ultra-sensitive mode[116] against the NCBI NR database and all the PhaeoExplorer proteins[18] to perform a phylostratigraphic analysis (*e*-value threshold of $10^{-5}$) and calculate the relative ages of each gene in each genome (Supplementary Table 13). Phylostratigraphy is a genetic statistical method developed to date the putative origin of all the genes contained in the genome of a target species by detecting homologues across species at different evolutionary distances (all the way from species within the same genus to species from different domains of life). Finding the most distant homologues of each gene can link them to their founder events (that is, the first instance where a gene homologue is found in the history of that lineage), allowing us to then determine their relative ages, coded as the taxonomic group where that gene is detected[35,36,129]. The gene age categories outside of the brown algae and *S. ischiensis* were based on the taxonomic classification of each species within the NCBI Taxonomy database[130], while the gene ages within the brown algae were manually assessed to reflect the evolutionary relationships obtained in the PhaeoExplorer maximum likelihood tree[18]. We performed Wilcoxon rank-sum tests in R (v.4.3.1)[131] to assess non-random differences in gene age distributions between pairs of chromosomes (Supplementary Table 14). We performed Benjamini–Hochberg corrections to the *P* values of the Wilcoxon rank-sum tests to control the FDR in the analysis[128]. The gene ages responsible for these differences were found by evaluating the standardized residuals using mosaic plots.

We used the interspecies $K_s$ values between pairs of species as a proxy for neutral mutation rates across six of the seven chromosome-level assemblies by using the most closely related genome assemblies available in the PhaeoExplorer database[18]. We used the same set of one-to-one orthologues detected between species pairs as for the orthologue-depletion test (Supplementary Table 16). However, the evolutionary distance between *D. dichotoma* and *H. paniculata* prevented us from calculating reliable interspecies $K_s$ values for this species since synonymous substitutions reached the point of saturation. The amino acid sequences of each pair of orthologues were aligned with MAFFT[117] and subsequently aligned into codons using pal2nal[118]. The interspecies $K_s$ values were calculated using the model in ref. 119 as implemented in KaKs_calculator (v.2.0)[120]. We also evaluated the difference in interspecies $K_s$ values between the autosomes and the sex chromosomes through FDR-corrected Wilcoxon rank-sum tests (Supplementary Table 17). We calculated the protein-coding density, the density of TEs and the taxonomic identity of these TEs within 100-kb non-overlaping windows across each chromosome using bedtools[86] (Supplementary Table 9). The differences in protein-coding space and repeat content between the autosomes and the sex chromosomes were also tested using FDR-corrected Wilcoxon rank-sum tests (Supplementary Tables 10 and 11). The differences in repeat density, percentage of unclassified repeats and gene ages across genomic compartments (SDR, PARs and autosomes) were tested using FDR-corrected permutation tests with 10,000 permutations. All genomic features were plotted using karyoploteR (v.1.20.3)[132].

### Gene expression analysis

We used kallisto (v.0.44.0)[133] to calculate gene expression levels using 31-base-pair-long *k*-mers and 1,000 bootstraps. Transcript abundances were then summed within genes using the tximport v.3.19 package[134] to obtain the expression level for each gene in transcripts per million (TPM). Differential expression analysis was done in the DESeq2 v.3.19 package[135] in R v.4.3.1, applying FC ≥ 2 and $p_{adj} < 0.05$ cut-offs. Sex-biased gene expression analysis in *Ectocarpus* sp. 7, *S. promiscuus*, *U. pinnatifida*, *D. herbacea* and *D. dichotoma* was performed between mature male and female gametophytes (gametophytes bearing reproductive structures). To discover genes with sporophyte-biased expression in *Ectocarpus* sp.7, *S. promiscuus*, *U.pinnatifida* and *D. dichotoma*, we first calculated the differential expression between male gametophytes and sporophytes, as well as female gametophytes and sporophytes. Genes that showed significant sporophyte-biased expression (FC ≥ 2, $p_{adj} < 0.05$) in both comparisons were considered sporophyte biased.

A total of 314.2 M RNA-seq reads from *F. vesiculosus* male, female and vegetative tissue were assembled de novo with rnaSPAdes[136] using *k*-mer values of 33 and 49. Assembly quality was assessed by (pseudo) mapping reads back onto the resulting assembly and retaining 'good' contigs as defined using TransRate (v.1.0.3)[137] with default settings. The resulting 159,108 contigs were aligned with BLASTx[82] against a database of Stramenopile proteins, and those with top hits against brown algae (Phaeophyceae) were retained as the final curated reference transcriptome (36,394 contigs, N50 = 1,770 bp). Transcript expression levels were determined by mapping the reads from all samples against the reference transcriptome using Bowtie2 (ref. 138) and the RSEM-EBSeq (v.1.3.3)[139] pipeline, and relative expression values were recorded as TPM. All samples used in the gene expression analysis can be found in Supplementary Table 1.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

The accession numbers and download links for all the chromosome-level genome assemblies and annotations that were generated and used in this study are available in Supplementary Table 1 and in the Edmond Repository[140] at https://doi.org/10.17617/3.OOWB2Y. The raw sequence reads for the Oxford Nanopore data, Hi-C libraries and RNA-seq libraries are available in the Sequence Read Archive under BioProject accession number PRJNA1059008. All genome assemblies and annotations are also accessible through the Phaeoexplorer database (https://phaeoexplorer.sb-roscoff.fr/) for comparative genomics

analyses. This paper does not report original data. Further information and requests for resources and reagents should be directed to and will be fulfilled by S.M.C. (susana.coelho@tuebingen.mpg.de).

## Code availability

This paper does not report original code.

## References


1. Bachtrog, D. et al. Sex determination: why so many ways of doing it? *PLoS Biol.* **12**, e1001899 (2014).
2. Beukeboom, L. W. & Perrin, N. *The Evolution of Sex Determination* (Oxford Univ. Press, 2014).
3. Vicoso, B. Molecular and evolutionary dynamics of animal sex-chromosome turnover. *Nat. Ecol. Evol.* **3**, 1632–1641 (2019).
4. Mignerot, L. & Coelho, S. M. The origin and evolution of the sexes: novel insights from a distant eukaryotic linage. *C. R. Biol.* **339**, 252–257 (2016).
5. Ma, W. & Rovatsos, M. Sex chromosome evolution: the remarkable diversity in the evolutionary rates and mechanisms. *J. Evol. Biol.* **35**, 1581–1588 (2022).
6. Umen, J. & Coelho, S. Algal sex determination and the evolution of anisogamy. *Annu. Rev. Microbiol.* **73**, 267–291 (2019).
7. Coelho, S. M. & Umen, J. Switching it up: algal insights into sexual transitions. *Plant Reprod.* **34**, 287–296 (2021).
8. Avia, K. et al. Genetic diversity in the UV sex chromosomes of the brown alga *Ectocarpus*. *Genes* **9**, 286 (2018).
9. Coelho, S. M., Mignerot, L. & Cock, J. M. Origin and evolution of sex-determination systems in the brown algae. *New Phytol.* **222**, 1751–1756 (2019).
10. Bull, J. J. Sex chromosomes in haploid dioecy: a unique contrast to Muller's theory for diploid dioecy. *Am. Nat.* **112**, 245–250 (1978).
11. Carey, S. B. et al. Gene-rich UV sex chromosomes harbor conserved regulators of sexual development. *Sci. Adv.* **7**, eabn2488 (2021).
12. Healey, A. L. et al. Newly identified sex chromosomes in the *Sphagnum* (peat moss) genome alter carbon sequestration and ecosystem dynamics. *Nat. Plants* **9**, 238–254 (2023).
13. Bowman, J. L. et al. Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* **171**, 287–304.e15 (2017).
14. Silva, A. T. et al. To dry perchance to live: insights from the genome of the desiccation-tolerant biocrust moss *Syntrichia caninervis*. *Plant J.* **105**, 1339–1356 (2021).
15. Bechteler, J. et al. Comprehensive phylogenomic time tree of bryophytes reveals deep relationships and uncovers gene incongruences in the last 500 million years of diversification. *Am. J. Bot.* **110**, e16249 (2023).
16. Heesch, S. et al. Evolution of life cycles and reproductive traits: insights from the brown algae. *J. Evol. Biol.* **34**, 992–1009 (2021).
17. Bringloe, T. T. et al. Phylogeny and evolution of the brown algae. *Crit. Rev. Plant Sci.* **39**, 281–321 (2020).
18. Denoeud, F. et al. Evolutionary genomics of the emergence of brown algae as key components of coastal ecosystems. *Cell* **187**, 6943–6965.e39 (2024).
19. Liu, P. et al. 3D chromatin maps of a brown alga reveal U/V sex chromosome spatial organization. *Nat. Commun.* **15**, 9590 (2024).
20. Shan, T. et al. First genome of the brown alga *Undaria pinnatifida*: chromosome-level assembly using PacBio and Hi-C technologies. *Front. Genet.* **11**, 140 (2020).
21. Choi, S.-W. et al. Ordovician origin and subsequent diversification of the brown algae. *Curr. Biol.* https://doi.org/10.1016/j.cub.2023.12.069 (2024).
22. Luthringer, R. et al. Repeated co-option of HMG-box genes for sex determination in brown algae and animals. *Science* **383**, eadk5466 (2024).
23. Simakov, O. et al. Deeply conserved synteny and the evolution of metazoan chromosomes. *Sci. Adv.* **8**, eabi5884 (2022).
24. Sun, Y., Svedberg, J., Hiltunen, M., Corcoran, P. & Johannesson, H. Large-scale suppression of recombination predates genomic rearrangements in *Neurospora tetrasperma*. *Nat. Commun.* **8**, 1140 (2017).
25. Ahmed, S. et al. A haploid system of sex determination in the brown alga *Ectocarpus* sp. *Curr. Biol.* **24**, 1945–1957 (2014).
26. Charlesworth, D. Plant contributions to our understanding of sex chromosome evolution. *New Phytol.* **208**, 52–65 (2015).
27. Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964–967 (1999).
28. Lipinska, A. P. et al. Multiple gene movements into and out of haploid sex chromosomes. *Genome Biol.* **18**, 104 (2017).
29. Charlesworth, D. The mysterious sex chromosomes of haploid plants. *Heredity* **129**, 17–21 (2022).
30. Bolwell, G. P., Callow, J. A., Callow, M. E. & Evans, L. V. Fertilization in brown algae: II. evidence for lectin-sensitive complementary receptors involved in gamete recognition in *Fucus serratus*. *J. Cell Sci.* **36**, 19–30 (1979).
31. Zhao, Z.-S., Leung, T., Manser, E. & Lim, L. Pheromone signalling in *Saccharomyces cerevisiae* requires the small GTP-binding protein Cdc42p and its activator *CDC24*. *Mol. Cell. Biol.* **15**, 5246–5257 (1995).
32. Ponnikas, S., Sigeman, H., Abbott, J. K. & Hansson, B. Why do sex chromosomes stop recombining? *Trends Genet.* **34**, 492–503 (2018).
33. Cossard, G. G. et al. Selection drives convergent gene expression changes during transitions to co-sexuality in haploid sexual systems. *Nat. Ecol. Evol.* **6**, 579–589 (2022).
34. Bachtrog, D. Accumulation of Spock and Worf, two novel non-LTR retrotransposons, on the neo-Y chromosome of *Drosophila miranda*. *Mol. Biol. Evol.* **20**, 173–181 (2003).
35. Domazet-Lošo, T., Brajković, J. & Tautz, D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* **23**, 533–539 (2007).
36. Barrera-Redondo, J., Lotharukpong, J. S., Drost, H.-G. & Coelho, S. M. Uncovering gene-family founder events during major evolutionary transitions in animals, plants and fungi using GenEra. *Genome Biol.* **24**, 54 (2023).
37. Luthringer, R. et al. The pseudoautosomal regions of the U/V sex chromosomes of the brown alga *Ectocarpus* exhibit unusual features. *Mol. Biol. Evol.* **32**, 2973–2985 (2015).
38. Kimura, M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**, 275–276 (1977).
39. King, J. L. & Jukes, T. H. Non-Darwinian evolution. *Science* **164**, 788–798 (1969).
40. Eyre-Walker, A. & Keightley, P. D. High genomic deleterious mutation rates in hominids. *Nature* **397**, 344–347 (1999).
41. Keightley, P. D. & Eyre-Walker, A. Deleterious mutations and the evolution of sex. *Science* **290**, 331–333 (2000).
42. Montgomery, S. A. et al. Chromatin organization in early land plants reveals an ancestral association between H3K27me3, transposons, and constitutive heterochromatin. *Curr. Biol.* **30**, 573–588.e7 (2020).
43. Loftus, B. J. et al. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* **307**, 1321–1324 (2005).
44. Clayton, M. N. Isogamy and a fucalean type of life history in the Antarctic brown alga *Ascoseira mirabilis* (Ascoseirales, Phaeophyta). *Bot. Mar.* **30**, 447–454 (1987).
45. Coyer, J. A., Peters, A. F., Hoarau, G., Stam, W. T. & Olsen, J. L. Hybridization of the marine seaweeds, *Fucus serratus and Fucus evanescens* (Heterokontophyta: Phaeophyceae) in a 100-year-old zone of secondary contact. *Proc. R. Soc. B* **269**, 1829–1834 (2002).

46. Hatchett, W. J. et al. Evolutionary dynamics of sex-biased gene expression in a young XY system: insights from the brown alga genus *Fucus*. *New Phytol.* **238**, 422–437 (2023).
47. Cánovas, F. G., Mota, C. F., Serrão, E. A. & Pearson, G. A. Driving south: a multi-gene phylogeny of the brown algal family Fucaceae reveals relationships and recent drivers of a marine radiation. *BMC Evol. Biol.* **11**, 371 (2011).
48. Lee, S. C., Ni, M., Li, W., Shertz, C. & Heitman, J. The evolution of sex: a perspective from the fungal kingdom. *Microbiol. Mol. Biol. Rev.* **74**, 298–340 (2010).
49. Duhamel, M., Hood, M. E., Rodríguez de la Vega, R. C. & Giraud, T. Dynamics of transposable element accumulation in the non-recombining regions of mating-type chromosomes in anther-smut fungi. *Nat. Commun.* **14**, 5692 (2023).
50. Dolgin, E. S. & Charlesworth, B. The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics* **178**, 2169–2177 (2008).
51. Gray, Y. H. M. It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. *Trends Genet.* **16**, 461–468 (2000).
52. Charlesworth, D. Evolution of recombination rates between sex chromosomes. *Phil. Trans. R. Soc. B* **372**, 20160456 (2017).
53. Jordan, C. Y. & Charlesworth, D. The potential for sexually antagonistic polymorphism in different genome regions. *Evolution* **66**, 505–516 (2012).
54. Luthringer, R. et al. Sexual dimorphism in the brown algae. *Perspect. Phycol.* **1**, 11–25 (2014).
55. Scharmann, M., Rebelo, A. G. & Pannell, J. R. High rates of evolution preceded shifts to sex-biased gene expression in *Leucadendron*, the most sexually dimorphic angiosperms. *Elife* **10**, e67485 (2021).
56. Charlesworth, B. Model for evolution of Y chromosomes and dosage compensation. *Proc. Natl Acad. Sci. USA* **75**, 5618–5622 (1978).
57. Lenormand, T., Fyon, F., Sun, E. & Roze, D. Sex chromosome degeneration by regulatory evolution. *Curr. Biol.* **30**, 3001–3006. e5 (2020).
58. Jeffries, D. L., Gerchen, J. F., Scharmann, M. & Pannell, J. R. A neutral model for the loss of recombination on sex chromosomes. *Phil. Trans. R. Soc. B* **376**, 20200096 (2021).
59. Jay, P., Tezenas, E., Véber, A. & Giraud, T. Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes. *PLoS Biol.* **20**, e3001698 (2022).
60. Coelho, S. M., Gueno, J., Lipinska, A. P., Cock, J. M. & Umen, J. G. UV chromosomes and haploid sexual systems. *Trends Plant Sci.* **23**, 794–807 (2018).
61. Peona, V. et al. Teaching transposon classification as a means to crowd source the curation of repeat annotation—a tardigrade perspective. *Mob. DNA* **15**, 10 (2024).
62. Ivics, Z. & Izsvák, Z. The expanding universe of transposon technologies for gene and cell engineering. *Mob. DNA* **1**, 25 (2010).
63. Gueno, J. et al. Chromatin landscape associated with sexual differentiation in a UV sex determination system. *Nucleic Acids Res.* **50**, 3307–3322 (2022).
64. Bourdareau, S. et al. Histone modifications during the life cycle of the brown alga *Ectocarpus*. *Genome Biol.* **22**, 12 (2021).
65. Makova, K. D. & Hardison, R. C. The effects of chromatin organization on variation in mutation rates in the genome. *Nat. Rev. Genet.* **16**, 213–223 (2015).
66. Charlesworth, B. The effects of deleterious mutations on evolution at linked sites. *Genetics* **190**, 5–22 (2012).
67. Lebherz, M. K., Fouks, B., Schmidt, J., Bornberg-Bauer, E. & Grandchamp, A. DNA transposons favor de novo transcript emergence through enrichment of transcription factor binding motifs. *Genome Biol. Evol.* https://doi.org/10.1093/gbe/evae134 (2024).
68. McDaniel, S. F. Divergent outcomes of genetic conflict on the UV sex chromosomes of *Marchantia polymorpha* and *Ceratodon purpureus*. *Curr. Opin. Genet. Dev.* **83**, 102129 (2023).
69. Iwasaki, M. et al. Identification of the sex-determining factor in the liverwort *Marchantia polymorpha* reveals unique evolution of sex chromosomes in a haploid system. *Curr. Biol.* **31**, 5522–5532.e7 (2021).
70. Ligrone, R., Duckett, J. G. & Renzaglia, K. S. The origin of the sporophyte shoot in land plants: a bryological perspective. *Ann. Bot.* **110**, 935–941 (2012).
71. McDaniel, S. F., Neubig, K. M., Payton, A. C., Quatrano, R. S. & Cove, D. J. Recent gene-capture on the UV sex chromosomes of the moss *Ceratodon purpureus*. *Evolution* https://doi.org/10.1111/evo.12165 (2013).
72. Vigneau, J. et al. Interactions between U and V sex chromosomes during the life cycle of *Ectocarpus*. *Development* **151**, dev202677 (2024).
73. Singh, S., Davies, K. M., Chagné, D. & Bowman, J. L. The fate of sex chromosomes during the evolution of monoicy from dioicy in liverworts. *Curr. Biol.* **33**, 3597–3609.e3 (2023).
74. Takahashi, K. et al. Reorganization of the ancestral sex-determining regions during the evolution of trioecy in *Pleodorina starrii*. *Commun. Biol.* **6**, 590 (2023).
75. Liesner, D. et al. Developmental pathways underlying sexual differentiation in the U/V sex chromosome system of giant kelp. *Dev. Cell* https://doi.org/10.1016/j.devcel.2024.12.022 (2025).
76. Yamamoto, K. et al. 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**, e2100712118 (2021).
77. Herpin, A. & Schartl, M. Plasticity of gene-regulatory networks controlling sex determination: of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO Rep.* **16**, 1260–1274 (2015).
78. Starr, R. C. & Zeikus, J. A. UTEX—the culture collection of algae at the University of Texas at Austin 1993 list of cultures. *J. Phycol.* **29**, 1–106 (1993).
79. Pearson, G., Lago-Leston, A., Valente, M. & Serrão, E. Simple and rapid RNA extraction from freeze-dried tissue of brown algae and seagrasses. *Eur. J. Phycol.* **41**, 97–104 (2006).
80. Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **20**, 129 (2019).
81. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
82. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
83. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).
84. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
85. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).
86. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
87. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
88. Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767–769 (2016).

89. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
90. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
91. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
92. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
93. Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639–1643 (2021).
94. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA* **117**, 9451–9457 (2020).
95. Alonge, M. et al. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* **23**, 258 (2022).
96. Carvalho, A. B. & Clark, A. G. Efficient identification of Y chromosome sequences in the human and *Drosophila* genomes. *Genome Res.* **23**, 1894–1907 (2013).
97. Vicoso, B., Emerson, J. J., Zektser, Y., Mahajan, S. & Bachtrog, D. Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. *PLoS Biol.* **11**, e1001643 (2013).
98. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764–770 (2011).
99. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **11**, e0163962 (2016).
100. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
101. Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2018).
102. Avia, K. et al. High-density genetic map and identification of QTLs for responses to temperature and salinity stresses in the model brown alga *Ectocarpus*. *Sci. Rep.* **7**, 43241 (2017).
103. Rochette, N. C., Rivera-Colón, A. G. & Catchen, J. M. Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol. Ecol.* **28**, 4737–4754 (2019).
104. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
105. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
106. Rastas, P. Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* **33**, 3726–3732 (2017).
107. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://doi.org/10.48550/arXiv.1303.3997 (2013).
108. Nursyifa, C., Brüniche-Olsen, A., Garcia-Erill, G., Heller, R. & Albrechtsen, A. Joint identification of sex and sex-linked scaffolds in non-model organisms using low depth sequencing data. *Mol. Ecol. Resour.* **22**, 458–467 (2022).
109. Malde, K., Skern, R. & Glover, K. A. Using sequencing coverage statistics to identify sex chromosomes in minke whales. Preprint at https://doi.org/10.48550/arXiv.1902.06654 (2019).
110. Smith, J. J. et al. The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nat. Genet.* **50**, 270–277 (2018).
111. Darolti, I. et al. Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. *Proc. Natl Acad. Sci. USA* **116**, 19031–19036 (2019).
112. Brelsford, A., Lavanchy, G., Sermier, R., Rausch, A. & Perrin, N. Identifying homomorphic sex chromosomes from wild-caught adults with limited genomic resources. *Mol. Ecol. Resour.* **17**, 752–759 (2017).
113. Toups, M. A., Rodrigues, N., Perrin, N. & Kirkpatrick, M. A reciprocal translocation radically reshapes sex-linked inheritance in the common frog. *Mol. Ecol.* **28**, 1877–1889 (2019).
114. Tang, H. et al. Synteny and collinearity in plant genomes. *Science* **320**, 486–488 (2008).
115. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
116. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).
117. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
118. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, 609–612 (2006).
119. Yang, Z. & Nielsen, R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**, 32–43 (2000).
120. Wang, D., Zhang, Y., Zhang, Z., Zhu, J. & Yu, J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics* **8**, 77–80 (2010).
121. Price, M. N., Dehal, P. S. & Arkin, A. P. Fasttree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).
122. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278 (2014).
123. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
124. Csűös, M. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* **26**, 1910–1912 (2010).
125. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
126. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
127. Chevenet, F., Brun, C., Bañuls, A.-L., Jacq, B. & Christen, R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **7**, 439 (2006).
128. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).
129. Lotharukpong, J. S. et al. A transcriptomic hourglass in brown algae. *Nature* **635**, 129–135 (2024).
130. Schoch, C. L. et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database* **2020**, baaa062 (2020).
131. R Core Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2021); https://www.r-project.org/
132. Gel, B. & Serra, E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**, 3088–3090 (2017).
133. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

134. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res.* **4**, 1521 (2015).
135. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
136. Bushmanova, E., Antipov, D., Lapidus, A. & Prjibelski, A. D. rnaSPAdes: a de novo transcriptome assembler and its application to RNA-seq data. *Gigascience* **8**, giz100 (2019).
137. Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M. & Kelly, S. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Res.* **26**, 1134–1144 (2016).
138. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
139. Leng, N. et al. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* **29**, 1035–1043 (2013).
140. Haas, F. B. Several brown algal chromosome level assemblies, V1. *Edmond* https://doi.org/10.17617/3.OOWB2Y (2025).
141. Cormier, A. et al. 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 (2017).

## Acknowledgements

This work was supported by the MPG, the CNRS, Sorbonne University, the ERC (grant no. 864038 and 638240 to S.M.C.), the France Génomique National infrastructure project Phaeoexplorer (ANR-10-INBS-09), the JSPS Overseas Research Fellowships (to M.H.), the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) (031A532B, 031A533A, 031A533B, 031A534A, 031A535A, 031A537A, 031A537B, 031A537C, 031A537D, 031A538A), the Investissements d'Avenir project Idealg (ANR-10-BTBR-04-01), the European BG-01 Blue Growth H2020 project Genialg (727892) and the ANR project Epicycle (ANR-19-CE20-0028-01). S.M.C. was supported by the Moore Foundation (GBMF11489) and the Bettencourt-Schueller Foundation. J.B.-R. was supported by a Humboldt Research Fellowship for postdoctoral researchers from the Alexander von Humboldt Foundation. We thank the members of the Phaeoexplorer consortium, in particular C. Jolivet, L. Mest and D. Scornet for assistance with algae cultures, C. Cruaud for help with sequencing libraries preparation, E. Corre and A. Le Bars for support with the Phaeoexplorer database and A. Couloux for the genome assemblies and annotations. We also thank the Roscoff Bioinformatics platform ABiMS (http://abims.sb-roscoff.fr), part of the Institut Français de Bioinformatique (ANR-11-INBS-0013) and Biogenouest network, for providing computing and storage resources.

## Author contributions

J.B.-R. and A.P.L. conducted investigation (equal) and formal analysis (equal), designed the methodology (equal), performed visualization (equal), wrote the original draft (equal), and reviewed and edited the manuscript (equal). P.L. conducted investigation (supporting) and formal analysis (supporting). E.D., G.C., O.G., K.B., M.H., R.J.C., K.A., G.L., E.A., D.L., R.L., O.G., S.H., Z.N., L.B.-G. and A.F.P. conducted investigation (supporting). G.H., J.-M.A,. G.P., P.W., F.D. and J.M.C. performed data curation (supporting) and data acquisition (supporting). F.B.H. conducted investigation (supporting), designed the methodology (equal), and performed data curation (equal) and formal analysis (supporting). S.M.C. conceptualized the project (lead), acquired funding (lead), administered and supervised the roject (lead), designed the methodology (equal), performed visualization (supporting), wrote the original draft (equal), and reviewed and edited the manuscript (lead).

## Funding

Open access funding provided by Max Planck Society.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at https://doi.org/10.1038/s41559-025-02838-w.

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41559-025-02838-w.

**Correspondence and requests for materials** should be addressed to Susana M. Coelho.

**Peer review information** *Nature Ecology & Evolution* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

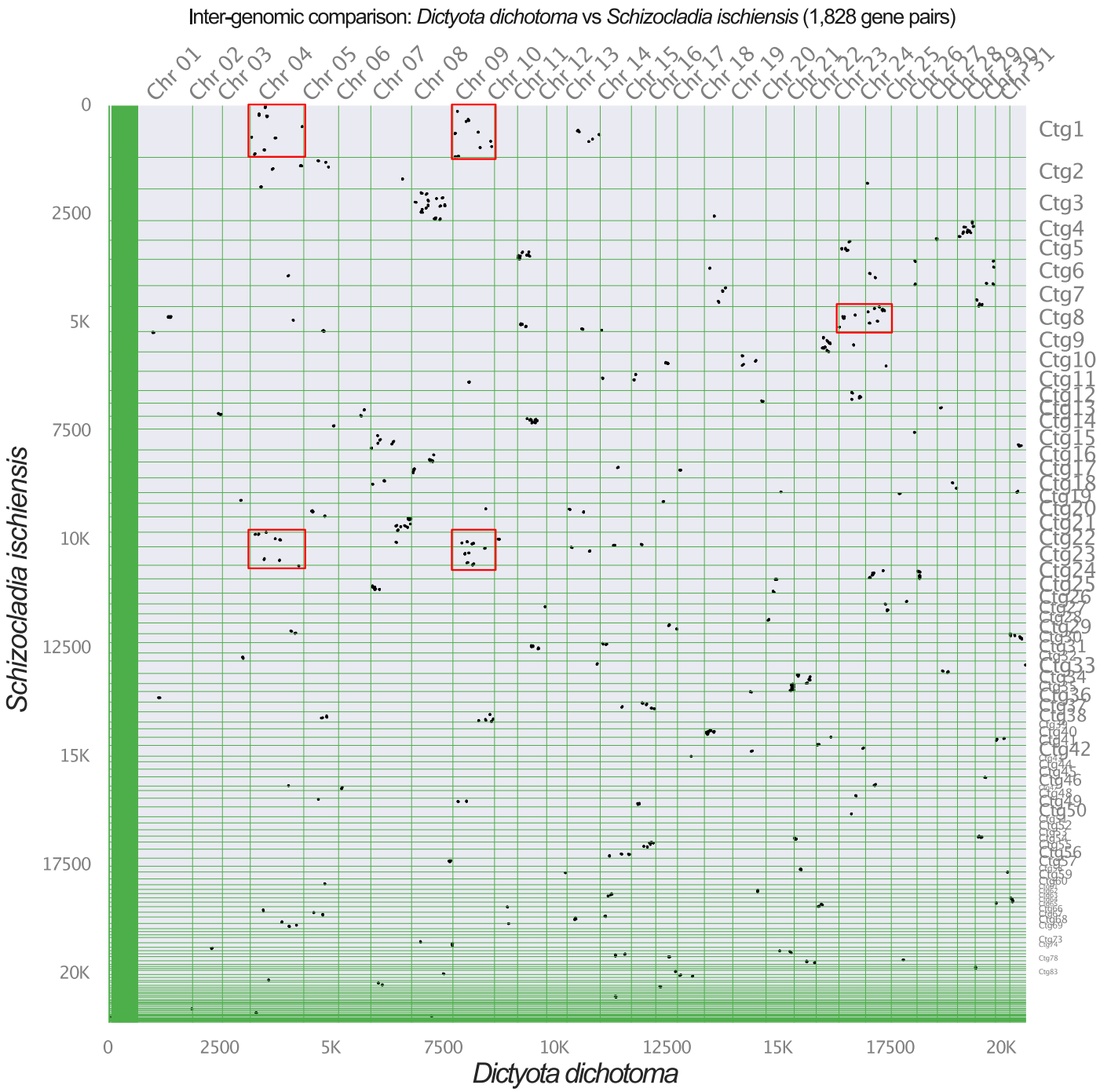
**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



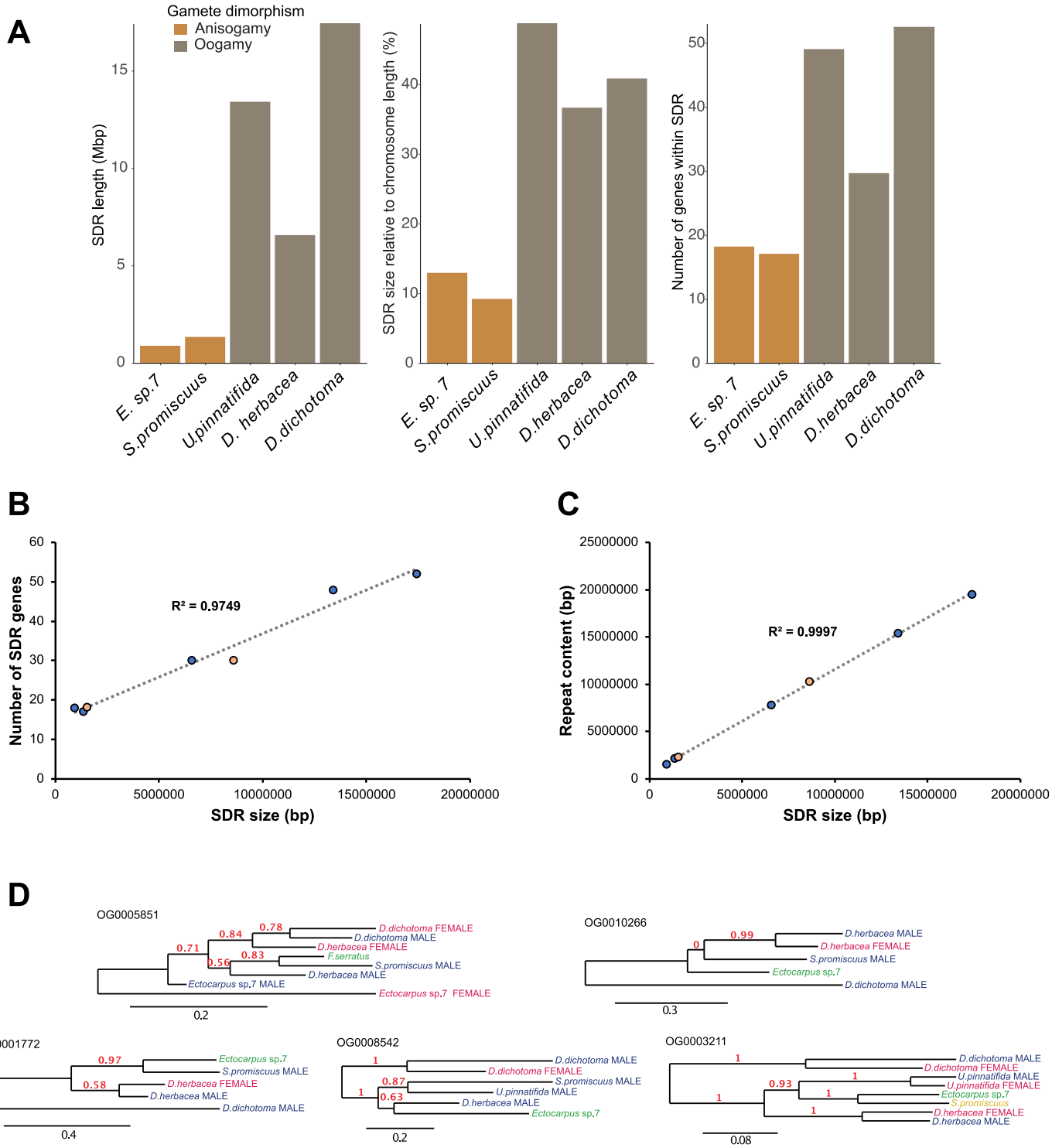
[1]Department of Algal Development and Evolution, Max Planck Institute for Biology Tübingen, Tübingen, Germany. [2]Sorbonne Université, CNRS, Integrative Biology of Marine Models Laboratory, Station Biologique de Roscoff, Roscoff, France. [3]INRAE, Université de Strasbourg, UMR SVQV, Colmar, France. [4]CNRS, Sorbonne Université, FR2424, ABiMS-IFB, Station Biologique, Roscoff, France. [5]Bezhin Rosko, Santeg, France. [6]Faculty of Biosciences and Aquaculture, Nord University, Bodø, Norway. [7]Universidade do Algarve, UALG·Centro de Ciências do Mar (CCMAR), Montenegro, Portugal. [8]Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France. [9]Present address: Departamento de Biotecnología y Bioquímica, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Unidad Irapuato, Irapuato, Mexico. [10]Present address: Research Center for Inland Seas, Kobe University, Nadaku, Kobe, Japan. [11]These authors contributed equally: Josué Barrera-Redondo, Agnieszka P. Lipinska. [12]These authors jointly supervised this work: France Denoeud, J. Mark Cock, Fabian B. Haas, Susana M. Coelho.
✉e-mail: susana.coelho@tuebingen.mpg.de

https://doi.org/10.1038/s41559-025-02838-w



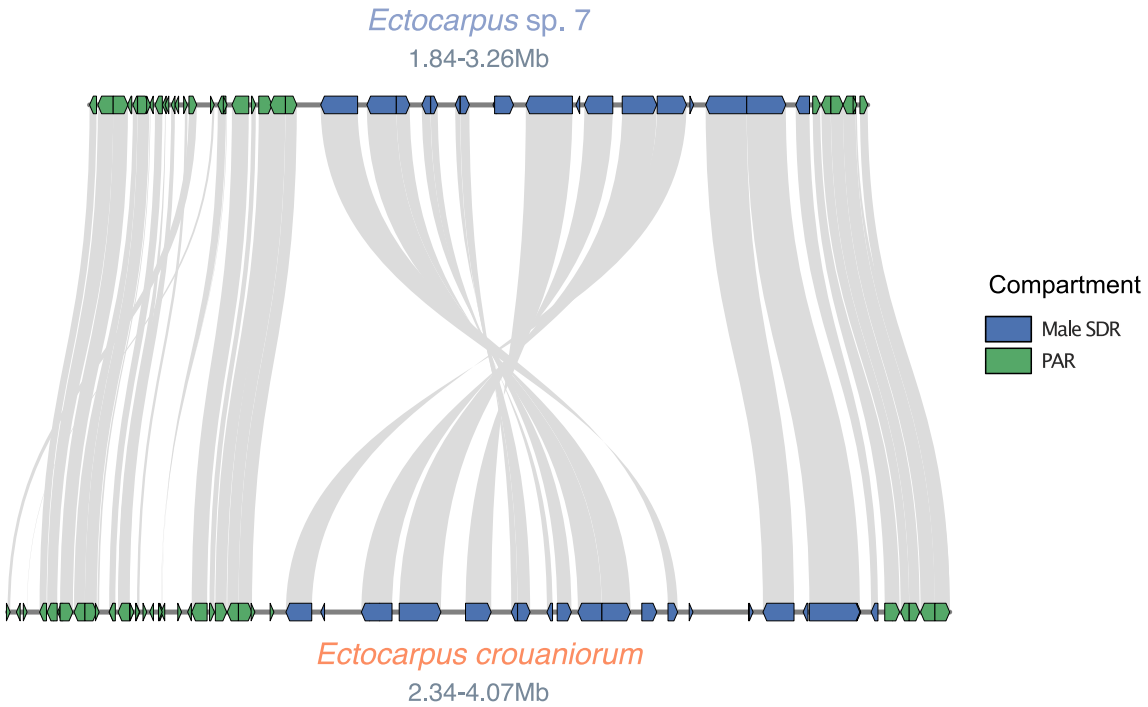

**Extended Data Fig. 1 | Macrosynteny plot between *Schizocladia ischiensis* and *Dictyota dichotoma* using 1,828 orthologs.** We highlight two fusion-with-mixing events (red squares) between chromosomes 4 and 9, and between chromosomes 23 and 24 in *D. dichotoma*.

https://doi.org/10.1038/s41559-025-02838-w



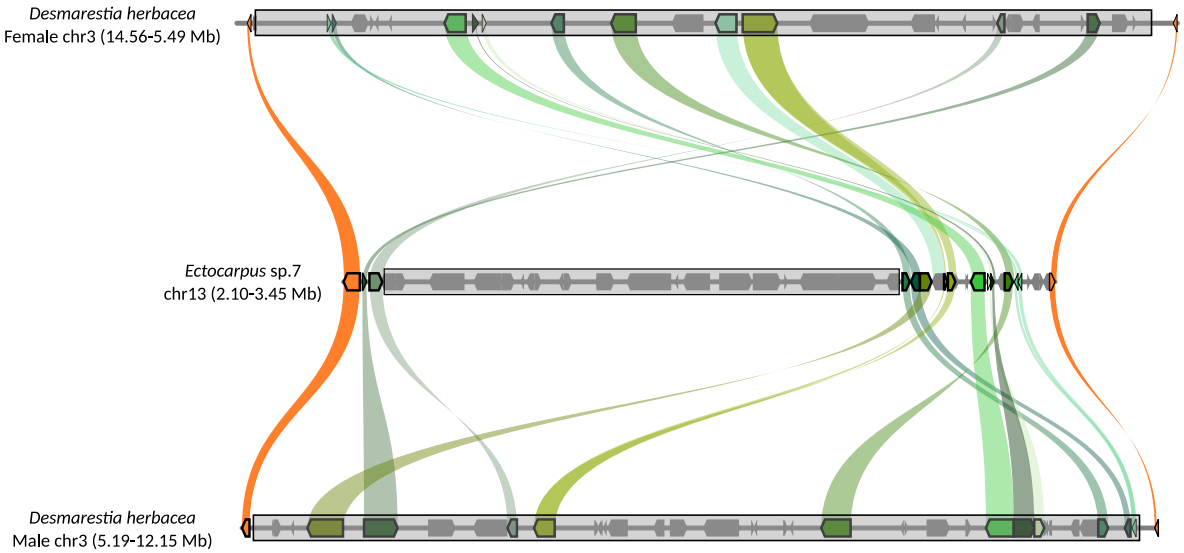

**Extended Data Fig. 2 | SDR size differences and detection of independently-acquired V-SDR genes across species.** (**a**) Differences in the size of the male SDR between brown algal species based on the total sequence length, the relative size of the SDR compared to the length of the V chromosome and the number of protein-coding genes retained within the SDR. The bars are colored accord-ing to the level of gamete dimorphism in each species (based on 16). (**b**) Correlation between the V (blue) and U (pink) SDR sizes and the SDR gene content across species. (**c**) Correlation between the V (blue) and U (pink) SDR sizes and the SDR repeat content across species. (**d**) Gene trees showing the independent acquisition of SDR gametologues across species that were previously interpreted as part of the ancestral male SDR genes.

https://doi.org/10.1038/s41559-025-02838-w



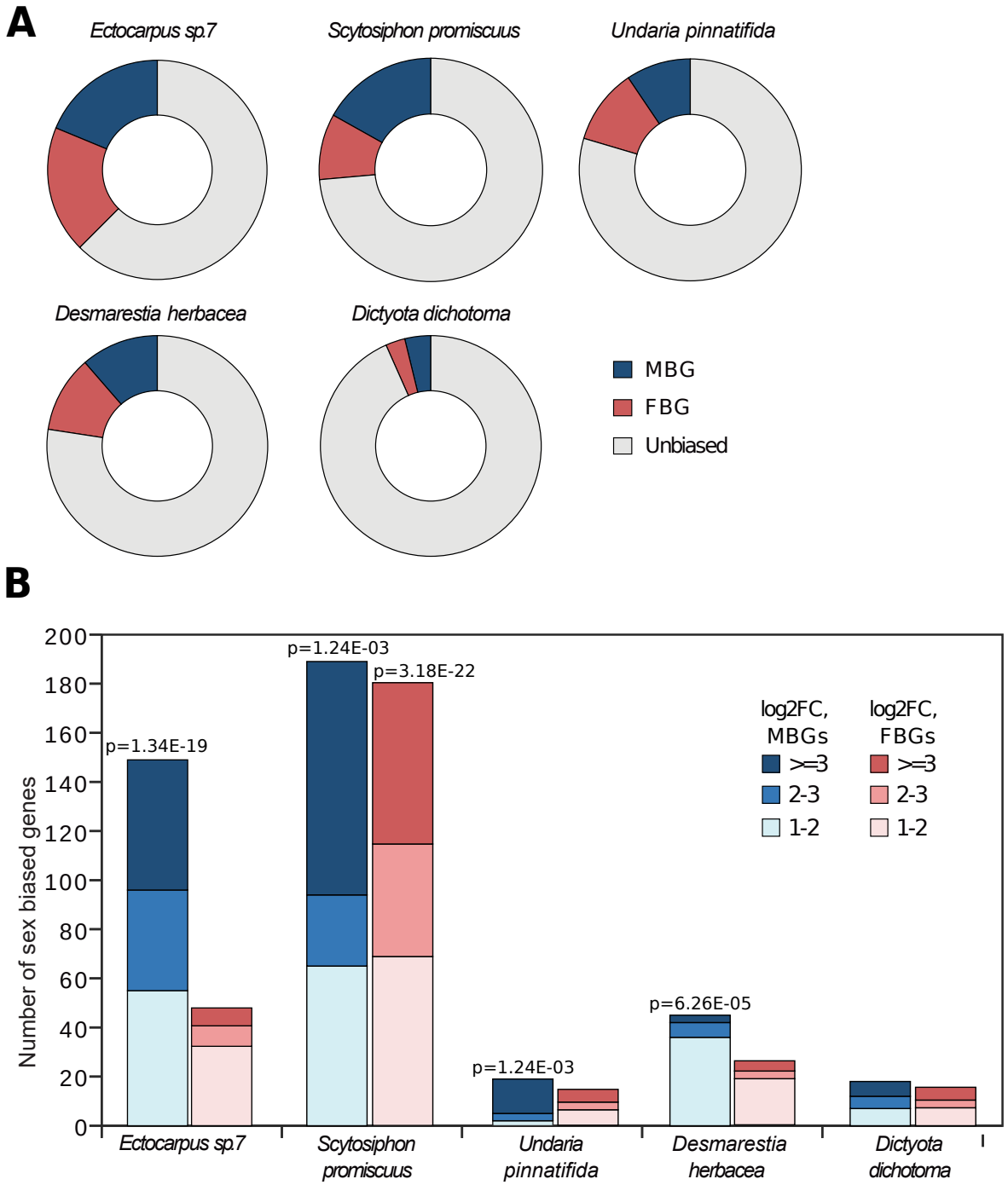

**Extended Data Fig. 3 | Male SDR synteny between *Ectocarpus* sp. 7 and *Ectocarpus crouaniorum*.** One of the species underwent a recent inversion event within the SDR. The arrows in the boxes represent the orientation of each gene within the chromosome.

https://doi.org/10.1038/s41559-025-02838-w



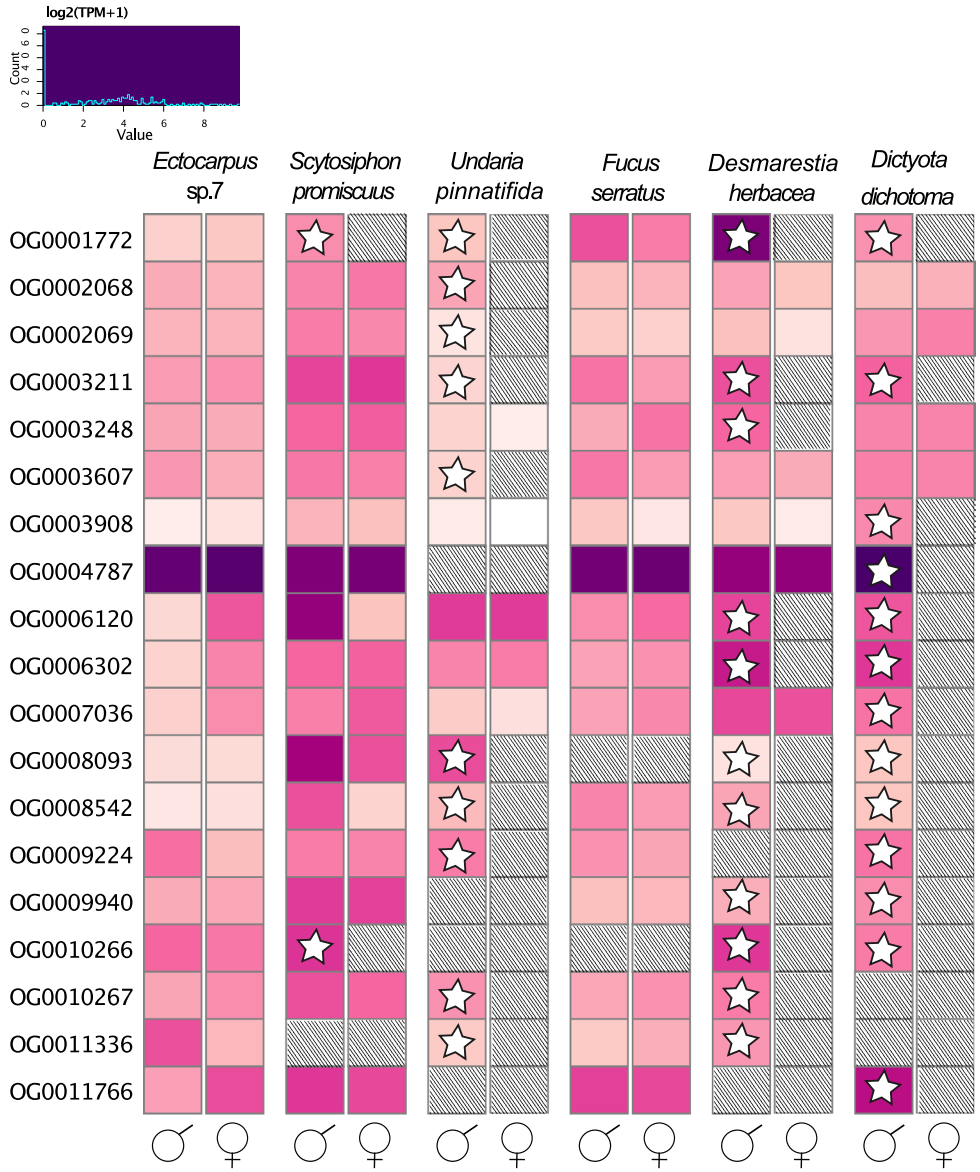

**Extended Data Fig. 4 | Synteny analysis plot illustrating the expansion of the *Desmarestia herbacea* U- and V-sex-determining regions (SDRs) into the surrounding pseudoautosomal region (PAR).** The *Ectocarpus* sp. 7 V-chromosome is shown in the middle as a reference, with the SDR regions outlined by grey boxes. Green lines trace syntenic relationships between *Ectocarpus* PAR genes and the recently-acquired *Desmarestia* SDR genes, with each gene pair represented in a distinct shade of green. This demonstrates that nearly all PAR genes from *Ectocarpus*, which have entered the expanded SDR in *Desmarestia*, are retained as gametologues. Orange lines highlight the PAR boundary genes in *Desmarestia*, which remain within the PAR of *Ectocarpus*.

https://doi.org/10.1038/s41559-025-02838-w



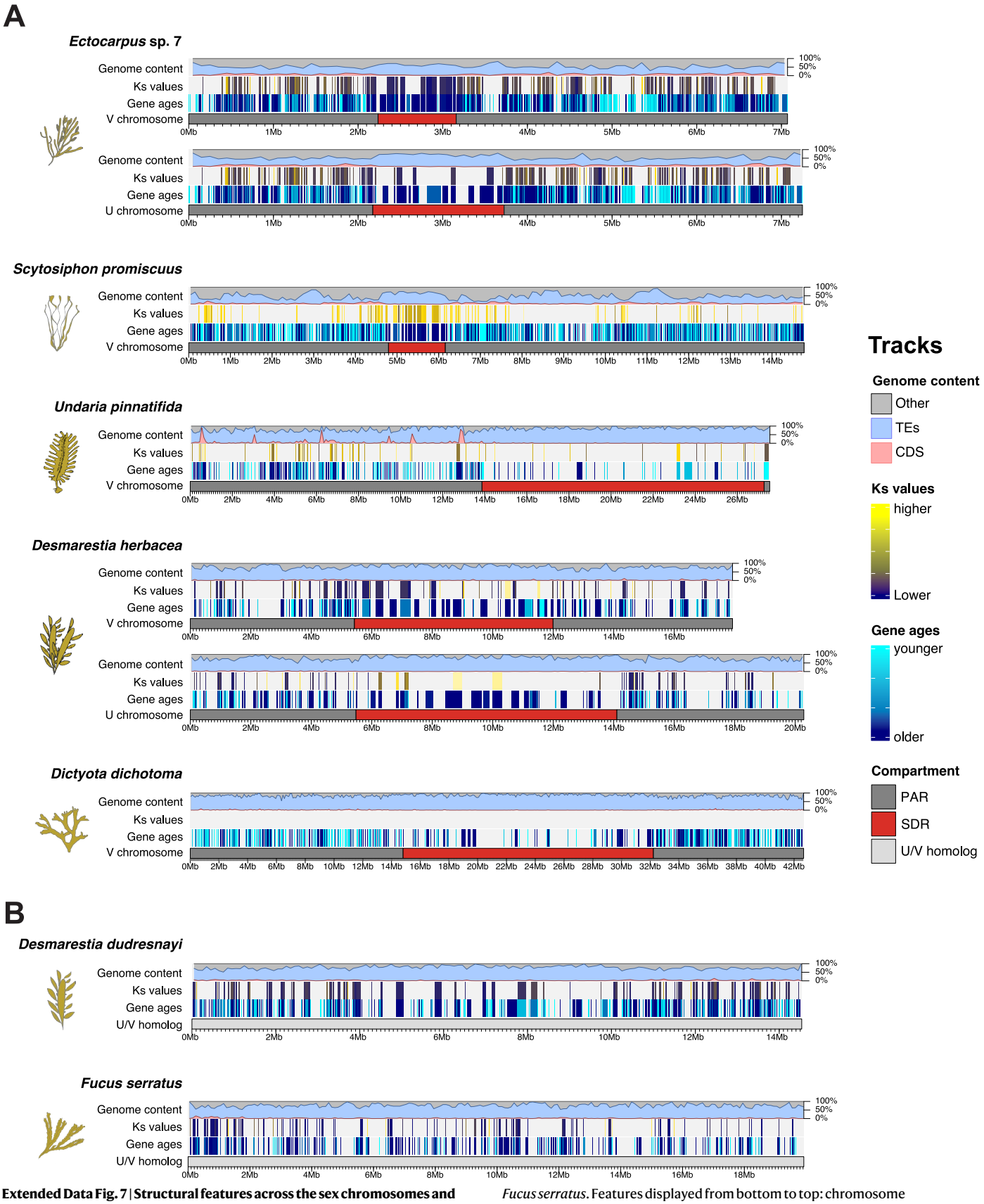

**Extended Data Fig. 5 | Sex-biased gene expression per dioicous species.** **a**) Proportion of sex biased genes in each of the five dioicous species. MBG: male-biased genes; FBG: female biased genes. (**b**) Number of sex-biased genes in the pseudoautosomal regions of sex chromosomes (U-V-SDRs excluded), male-biased genes are shown in blue and female-biased genes in red. Exact p-values above the bars mark significant enrichment of the sex-biased genes on the PAR (Chi-square test).

https://doi.org/10.1038/s41559-025-02838-w



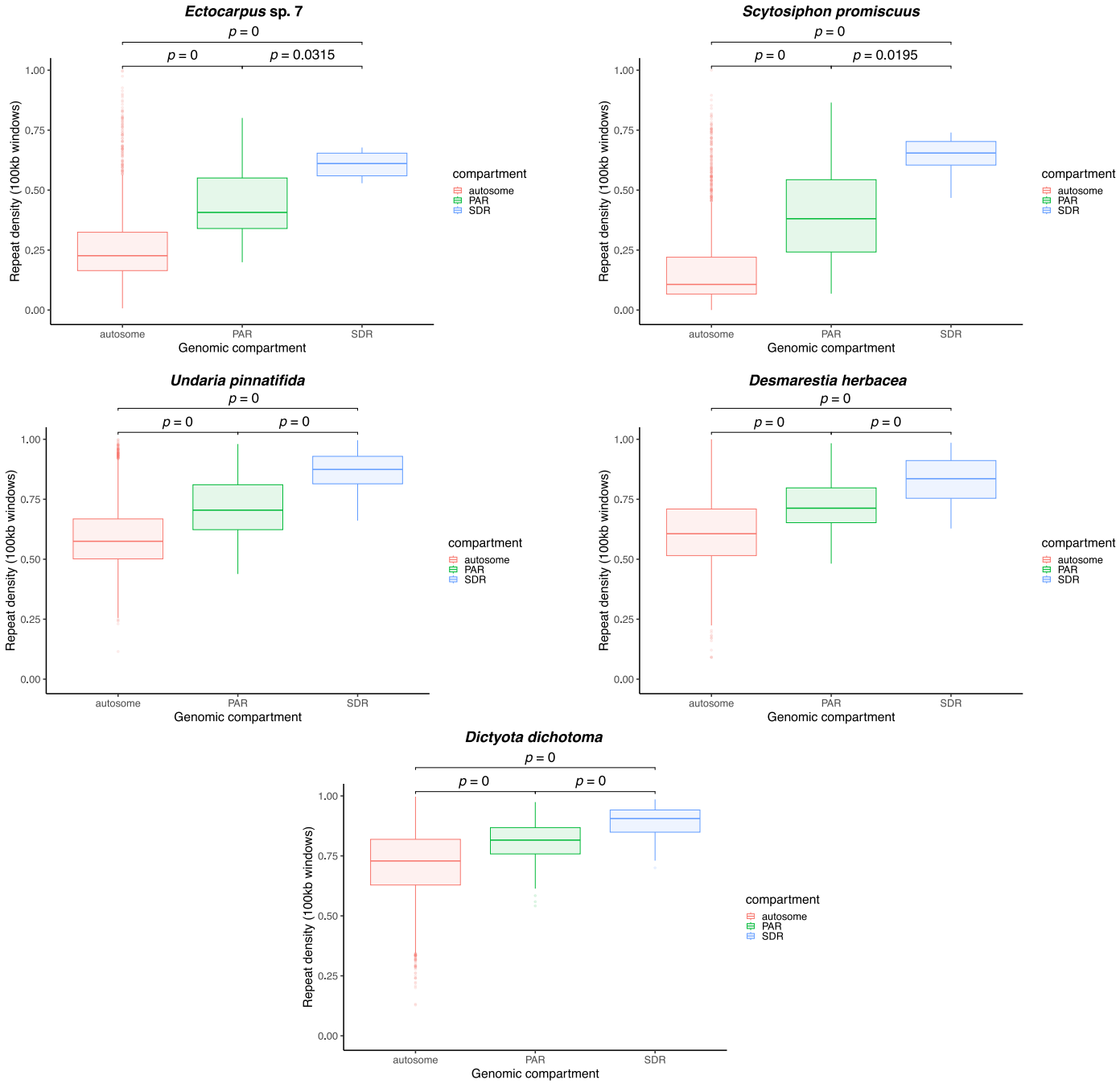

**Extended Data Fig. 6 | Expression of genes (log2(TPM + 1)) that entered the SDR independently in different species.** Expression is measured in mature male and female gametophytes, hashing marks missing orthologues, stars inside the cells indicate that the gene is inside the male non-recombining region (V-SDR). Orthogroups containing orthologues in less than three species or with multicopy genes were excluded from this analysis. M: male; F: female.

https://doi.org/10.1038/s41559-025-02838-w



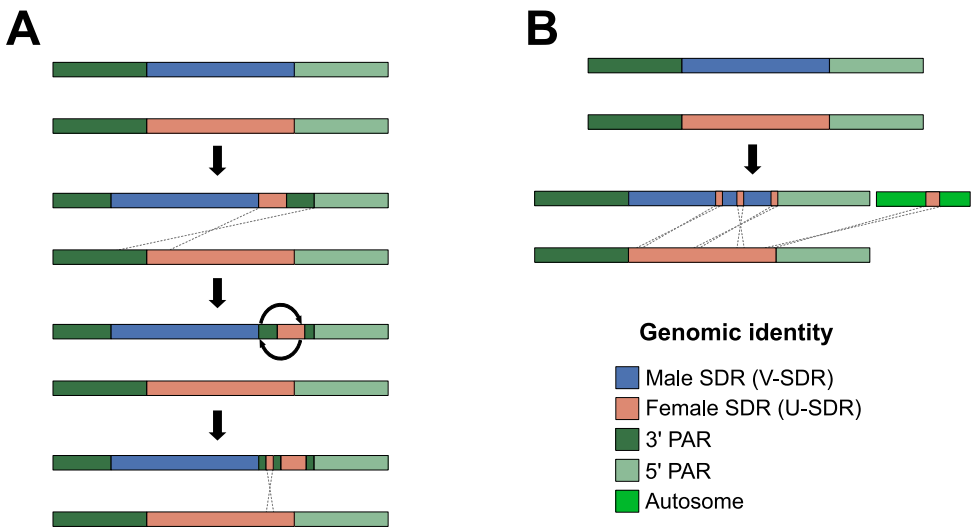

**Extended Data Fig. 7 | Structural features across the sex chromosomes and U/V-homologues of brown algae.** (**a**) V and U chromosomes of *Ectocarpus* sp. 7, *Scytosiphon promiscuus*, *Undaria pinnatifida*, *Desmarestia herbacea* and *Dictyota dichotoma*. (**b**) U/V-homologues of *Desmarestia dudresnayi* and *Fucus serratus*. Features displayed from bottom to top: chromosome compartments (PARs, SDR, U/V-homologue); relative gene ages, inter-species *Ks* values, and proportion of coding (CDS, red) and repeat (TEs, blue) density.

https://doi.org/10.1038/s41559-025-02838-w




**Extended Data Fig. 8 | Accumulation of repetitive elements in the V-SDRs and PARs of five dioicous species.** Statistically significant differences in median values of repeat density (center line) were assessed using FDR-corrected permutation tests.

https://doi.org/10.1038/s41559-025-02838-w




**Extended Data Fig. 9 | Proposed scenarios for the transition from dioicy to monoicy in *Chordaria linearis* and *Desmarestia dudresnayi*.** (**a**) The ancestor of *Chordaria linearis* likely underwent an initial translocation event from the U chromosome to the V chromosome, inserting part of the U-SDR and a piece of the 3′ PAR towards the 5′ end of the V-SDR potentially through an ectopic recombination event. A subsequent inversion within this translocation spread the 3′PAR genes to both sides of the U-SDR insertion. Finally, a second translocation led to the insertion of an additional piece of the U-SDR within the 3′ PAR translocation. (**b**) The ancestor of *Desmarestia dudresnayi* underwent three translocations of U-SDR genes into the V-SDR. Additionally, a fourth translocation event happened between the U-SDR and an autosome (chr_04).

https://doi.org/10.1038/s41559-025-02838-w

## Extended Data Table 1 | General characteristics of the genomes and sex chromosomes in representative brown algal species and an outgroup (*Schizocladia ischiensis*)

| Sexual system | Dioicous (V) | | | | | | Dioicous (U) | | Dioecious | Monoicous | | Unknown |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Species | *E. sp. 7* | *E.crou* | *S.prom* | *U. pinn* | *D. herb* | *D. dich* | *E. sp. 7* | *D. herb* | *F. serr* | *C. lin* | *D. dud* | *S. isch* |
| Genome size (Mbp) | 200.170 | 218.478 | 193.199 | 511.280 | 430.876 | 851.153 | 197.371 | 484.711 | 1 091.539 | 214.613 | 425.034 | 194.512 |
| Chrom-level scaffolds | 27 | 27 | 27 | 30 | 32 | 31 | 27 | 32 | 33 | NA | 32 | NA |
| Total n. scaffolds | 33 | 175 | 451 | 114 | 935 | 1467 | 46 | 1756 | 3,805 | 217 | 823 | 130 |
| N50 (bp) | 6 893 608 | 7 597 912 | 6 781 292 | 16 510 065 | 13 135 153 | 26 426 000 | 6 914 137 | 12 441 740 | 17 863 197 | 2 249 057 | 13 227 088 | 2 524 267 |
| Sex chrom. or U/V-homolog | chr_13 | chr_13 | chr_13 | HiC_scaffold_23 | chr_03 | chr_02 | chr_13 (Ec25_SDR_F) | chr_03 | LG15 | C-linearis_contig12 | chr_03 | NA |
| Sex chrom. or U/V-homolog size (bp) | 7 072 209 | 9 658 235 | 14 770 496 | 27 543 478 | 17 928 803 | 42 672 188 | 7 248 464 | 20 286 227 | 19 924 529 | 3 613 932 | 14 536 050 | NA |
| SDR length (bp) | 923 344 | 1 084 112 | 1 363 928 | 13 409 094 | 6 568 004 | 17 412 526 | 1 551 053 | 8 626 479 | NA | NA | NA | NA |
| N. genes in SDR | 18 | 18 | 17 | 48 | 30 (+ 20 viral-derived genes) | 52 | 18 | 30 | NA | NA | NA | NA |
| N. genes PAR | 421 | 519 | 904 | 306 | 229 | 451 | 421 | 229 | NA | NA | NA | NA |
| Gamy | anisogamous | anisogamous | anisogamous | oogamous | oogamous | oogamous | anisogamous | oogamous | NA | NA | NA | NA |
| CDS density SDR | 2.96% | NA | 2.31% | 0.51% | 0.99% | 0.28% | 1.75% | 0.81% | NA | NA | NA | NA |
| CDS density PAR or U/V-homolog | 8.50% | NA | 6.69% | 7.35% | 1.98% | 1.70% | 8.90% | 1.98% | 2.07% | NA | 2.83% | NA |
| CDS density autosomes | 14.97% | NA | 15.94% | 13.51% | 5.17% | 2.95% | 14.97% | 5.17% | 2.28% | NA | 9.27% | NA |
| Repeat density SDR | 60.64% | NA | 64.29% | 86.94% | 83.82% | 89.19% | 68.88% | 84.53% | NA | NA | NA | NA |
| Repeat density PAR or U/V homolog | 44.31% | NA | 40.19% | 72.51% | 72.92% | 80.88% | 41.46% | 71.76% | 74.26% | NA | 74.68% | NA |
| Repeat density autosomes | 30.12% | NA | 18.92% | 59.24% | 61.53% | 72.10% | 29.38% | 60.97% | 75.16% | NA | 57.63% | NA |
| Genome source | Liu et al.[18] | This study | This study | Shan et al.[19] | This study | This study | Liu et al.[18] | This study | This study | Denoeud et al.[17] | This study | Denoeud et al.[17] |

Note: Note that *Fucus serratus* (F. serr) has a chromosome (LG15) that is homologous to the ancestral U/V sex chromosome because it contains several genes present in the ancestral U/V-SDR. However, none of these genes are sex-linked in *F. serratus* (see text for details). Note that the dioecious species *F. serratus* has male and female (diploid) sexes and no gametophyte generation, that is, an animal-like life cycle. The CDS and repeat densities of the U/V-homologues were placed in the same rows as the PAR values of the dioicous species.

nature portfolio

Corresponding author(s): Susana Coelho

Last updated by author(s): 22/06/2025

# Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | details are provided in the method section of the manuscript |
| Data analysis | details on data analysis are provided in the methods section of the manuscript |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data and accession codes are provided in the supplemental tables of the manuscript. There are no restrictions on data availability

## Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

| | |
|---|---|
| Reporting on sex and gender | *Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used.*<br>*Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.*<br>*Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.* |
| Reporting on race, ethnicity, or other socially relevant groupings | *Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).*<br>*Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)*<br>*Please provide details about how you controlled for confounding variables in your analyses.* |
| Population characteristics | *Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."* |
| Recruitment | *Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.* |
| Ethics oversight | *Identify the organization(s) that approved the study protocol.* |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Study description | We identified the sex chromosomes and sex determining regions of several brown algal species represenitng the full phylogeny and level of morphological complexity across the lineage. |
| Research sample | brown algal lines |
| Sampling strategy | samples for each brown algal species used were grown in the laboratory in separate petri dishes as haploids. Sampeles were obtained from RCC (culture collection of Roscoff) and all details of codes and origin are provided in the manuscript. |
| Data collection | na |
| Timing and spatial scale | na |
| Data exclusions | no data was excluded |
| Reproducibility | for expression analysis the experiments were performed at least in triplicate |
| Randomization | samples were grown in the culture room in a randomized fashion. |
| Blinding | n.a |

Did the study involve field work? ☐ Yes ☒ No

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |
| ☒ | ☐ Plants |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Plants

Seed stocks

*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*

Novel plant genotypes

*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*

Authentication

*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

# Rewiring of chromatin regulation underlies the evolution of brown algal multicellularity

Jeromine Vigneau[1*], Jaruwatana Sodai Lotharukpong[1*], Pengfei Liu[1], Remy Luthringer[1], Bérangère Lombard[2], Damarys Loew[2], Fabian B. Haas[1], Michael Borg[1†], Susana M Coelho[1†]

[1]Department of Algal Development and Evolution, Max Planck Institute for Biology, Max-Planck-Ring 5, 72076 Tübingen, Germany

[2]Institut Curie, PSL Research University, CurieCoreTech Mass Spectrometry Proteomics, Paris, France

[*]These authors contributed equally to this work.

[†]susana.coelho@tuebingen.mpg.de, michael.borg@tuebingen.mpg.de

## Abstract

Chromatin structure plays a central role in regulating transcription, genome stability, and epigenetic inheritance in eukaryotes. Much of our understanding of chromatin architecture and histone post-translational modifications (hPTMs) comes from a narrow set of animal and plant models, but emerging data from non-model lineages are challenging canonical views of how chromatin functions across the tree of life. Brown algae are complex multicellular eukaryotes that provide a unique perspective on chromatin evolution given their independent origin of complex multicellularity. Here, we compile the chromatin toolkit of brown algae and show that canonical silencing systems involving DNA cytosine methylation and PRC2-mediated H3K27 methylation were lost early in their evolution. By generating hPTM profiles from diverse brown algal clades, we resolve the nature and regulatory roles of chromatin states in this lineage and show how H3K79 methylation emerged and diversified as a repressive system. We further uncover sex-specific reconfigurations in species with varying degrees of sexual dimorphism and reconstruct the ancestral regulatory landscape that likely preceded the emergence of brown algae. Together, our findings illuminate the dynamic evolution of chromatin regulation in a distinct multicellular lineage and challenge assumptions about the universality of chromatin-based mechanisms across eukaryotes.

## Introduction

In eukaryotes, chromatin regulates access to the genome by packaging DNA into an organised structure that modulates transcription and other DNA-based processes. This organization is defined by chromatin states, which consist of specific combinations of DNA and/or histone post-translational modifications (hPTMs), DNA-binding proteins, and 3D structural features[1–3]. Chromatin states are often dynamic and responsive to developmental and environmental cues, and is some cases can be stably maintained though epigenetic inheritance[4]. In addition to gene regulation, chromatin plays a critical role in safeguarding genome integrity by repressing transposable elements and other invasive mobile elements[5].

Most of our understanding of epigenetic and chromatin-based regulation has largely been informed by the discovery and analysis of hPTMs in a limited range of plant and animal models[6]. While many hPTMs are evolutionarily conserved across eukaryotic lineages and trace back to the last eukaryotic common ancestor (LECA), recent work on non-model lineages is

revealing a surprising diversity of chromatin systems and challenges classical views about the conserved function of hPTMs[7]. For example, the brown alga *Ectocarpus* lacks canonical hPTMs such as H3K9 and H3K27 methylation, while H3K79 methylation appears to play a repressive role rather than associating with active transcription as in yeast and metazoans, highlighting a striking divergence in chromatin regulation[8–11]. The brown algae represent a distinct branch of complex multicellular eukaryotes, separated from animals and plants, that have evolved within the last 450 million years and display a broad diversity in genome size, morphology, and sexual systems[12–14]. With the recent development of genomic tools for this clade, including over 65 genome assemblies of which several are at a chromosome scale, as well as genetic tools, brown algae have emerged as attractive model organisms for comparative studies. Notably, their genomes have remained largely syntenic[15], which facilitates cross-species comparative genomic studies.

Here, we investigated the diversity of chromatin landscapes across major brown algal lineages that encompasses the broad morphological complexity and variation in sexual systems within the group, along with an outgroup species for comparison. We combined genome-wide chromatin modification maps with gene expression data and various genomic features to functionally interpret chromatin states across brown algae. Our analyses reveal several key insights into the evolution of chromatin landscapes in this group. First, the emergence of the brown algal lineage was accompanied by the loss of both PRC2-mediated repression and DNA methylation, representing a major shift in epigenetic regulation during early brown algal evolution. Second, activation-associated hPTMs are highly conserved across species, suggesting functional constraint. In contrast, the repressive role of H3K79 methylation does not appear to be ancestral but likely evolved prior to the divergence of the Ectocarpales and Laminariales. Third, sex-specific chromatin state differences between male and female gametophytes seems independent from the degree of sexual dimorphism observed at the organismal level, suggesting that morphological sex differentiation may be driven by chromatin reconfiguration at a limited number of high-order effector loci. Finally, by examining chromatin and DNA methylation in an outgroup species, we trace the transition from an ancestral epigenetic landscape to the distinct architecture observed in brown algae.

## Results

### Evolution of chromatin and epigenetic-related genes

The recent availability of several high-quality brown algal genomes allowed us to investigate the conservation and evolution of chromatin-associated proteins across the clade[12,15]. Using BLAST searches and orthology-based approaches, we screened for homologs of known chromatin-related proteins and revealed a conserved and distinct repertoire across brown algae (**Fig. 1A**). Consistent with previous reports, we observed a complete absence of MET1 orthologs and other DNA methyltransferases across the clade. MET1 orthologs were only identified in the closest non-Phaeophycean relative *Schizocladia ischiensis*, suggesting that the lineage-specific loss of DNA methyltransferases occurred early in brown algal evolution. Similarly, we found that EED and SUZ12 subunits specific to the Polycomb Repressive Complex 2 (PRC2) are absent from all surveyed brown algal species, while the modular MSI1

subunit common to other chromatin complexes was present. EED and SUZ12 orthologs are also absent in *S. ischiensis* as well as other closely-related Ochrophytina species, suggesting that the loss of PRC2 may have preceded the emergence of the Phaeophyceae. The loss of PRC2 was also reflected in the absence of PRC1 homologs, which forms a distinct repressive complex in animals and plants[16]. SET domain proteins are abundant across the clade, which included homologs of Trithorax-related histone methyltransferases and two brown algal-specific families of SET domain proteins. Interestingly, orthologs of DOT1, which is responsible for histone H3 lysine 79 (H3K79) methylation in yeast and animals[17], are greatly expanded among brown algae and have diversified into at least five distinct families (**Fig. S1**), highlighting a Phaeophyceae-specific adaptation in chromatin regulation. These observations show that the emergence of the brown algal lineage is marked by the concurrent loss of epigenetic control via PRC2 and DNA methylation and an expansion of DOT1 histone methyltransferases, suggesting a major shift in gene regulatory strategies in this clade.

To investigate the diversity of chromatin landscapes in brown algae, we selected a set of representative species spanning the phylogenetic breadth of the group, chosen to reflect variation in morphological complexity, sexual systems, and reproductive strategies (**Fig. 1B-C**). To enable comparisons between male and female gametophytes within the same genetic background, we used sibling samples where possible, thereby minimizing genetic variability unrelated to sex (**Table S1**). Our sampling included *Desmarestia dudresnayi* that recently transitioned to co-sexuality (monoicy), alongside its closest relative *Desmarestia herbacea,* a dioicous species with separate male and female individuals. We also included *Undaria pinnatifida*, a kelp species characterized by a highly complex morphology, high sexual dimorphism, and a large genome with expanded UV sex chromosomes[15]. This contrasts with *Ectocarpus* sp.7 and *Scytosciphon promiscuus*, which possess smaller sex-linked regions on their UV sex chromosomes and have low-to-medium sexual dimorphism, respectively[15,18,19] . Finally, we included the filamentous chromista *S. ischiensis* as a representative outgroup species from the closest diverging lineage outside of the brown algae[20].

We first performed mass spectrometry of histone preparations to detect hPTMs across the five representative species, as done previously in *Ectocarpus*[9] (**Fig. 1D; Table S2**). Overall, hPTMs were detected with minimal species-specific differences. Histones that we identified close to known H2A.Z variants (H2A.Z-like) carry a heavily acetylated tails are present in *Ectocarpus*, *U. pinnatifida*, *D. herbacea* and *S. ischiensis* but not retrieved in *S. promiscuus* and *D. dudresnayi*. H3K9me1 was detected in *D. herbacea* and *S. ischiensis*, and H3K9me3 in *U. pinnatifida*, but neither mark was observed in the other species (**Fig. 1D; Table S2**). Importantly, methylated forms of histone H3K27 were absent from all five brown algal species analysed, consistent with the lack of PRC2 subunits encoded in their genomes (**Fig. 1A, D; Table S2**). Taken together, this data indicates that the lack of canonical epigenetic regulation via DNA and H3K27 methylation is a general feature of brown algae.

In earlier studies focusing on chromatin landscapes in the filamentous brown alga *Ectocarpus*, we identified H3K4me3, H3K9ac and H3K36me3 as hPTMs of active chromatin, whereas H3K79me2 and H4K20me3 were more strongly linked to transcriptional repression [8,9]. Aside for H3K4me3 and H4K20me3, which were technically challenging to isolate in our mass

spectrometry runs, we were able to detect most of the other hPTMs in each species (**Table S2**). We further validated the presence of H3K4me3, H3K9ac and H3K36me3 in each species using immunoblotting although H3K79me2 was only detectable in *U. pinnatifida* (**Fig. 1D; Fig. S2** for full length uncropped blots**)**. We speculate that H3K79me2 is likely present at low levels and is challenging to detect with immunoblotting, which was further supported by similarly low signals in mouse histone extracts (**Fig. S2**). Building on their conservation across the clade and prior characterisation in *Ectocarpus*, we focused on this set of five hPTMs to investigate the evolutionary dynamics of chromatin landscapes in brown algae.

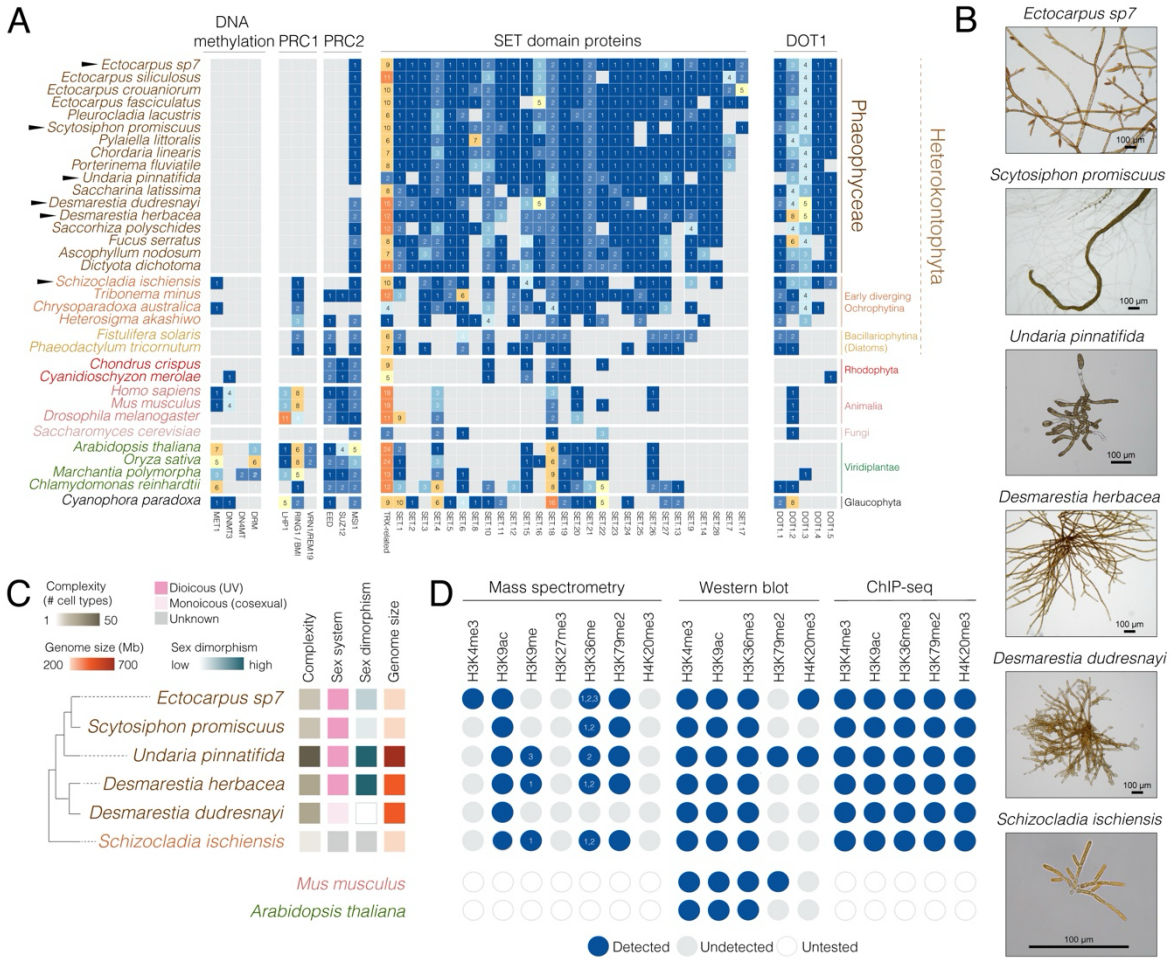



**Figure 1**. Evolution of chromatin-related and epigenetic pathways in brown algae. (**A**) Census of chromatin proteins across the brown algae, their closest relatives, and other major eukaryotic lineages. The species focused on in this study are indicated by a black triangle. (**B**) Representative images of the species used in this study. (**C**) Schematic diagram summarising the phylogenetic relationship of the six species studied alongside biological features that distinguish them, namely their overall morphological complexity, sexual system, sexual dimorphism and relative genome size. (**D**) Summary of the major hPTMs detected in the panel of six species using mass spectrometry, western blot and/or ChIP-seq. *Ectocarpus* mass spectrometry data was produced in a previous study[8].

## The chromatin landscape of diverse brown macroalgae

Closely related male and female gametophyte (haploid) lines were used to generate sex-specific

ChIP-seq profiles for the five histone PTMs in species with separate sexes (*Ectocarpus*, *S. promiscuus*, *U. pinnatifida*, *D. herbacea*) (**Table S1**; see Materials and Methods). Monoicous *D. dudrenayi* gametophytes and vegetative tissue from *S. ischiensis* were similarly profiled (**Fig. 2A**, **Fig. S3A**). For each species, we profiled at least 400 clonal individuals per male, female, monoicous or vegetative replicate, then merged ChIP-seq replicates after confirming high reproducibility (**Fig. S4; Table S4**).

Metaplot analyses revealed that H3K4me3 and H3K9ac were highly enriched at transcription start sites (TSSs) and positively correlated with gene expression across all brown algal species (**Fig. 2A; Fig. S5)**. H3K36me3 was consistently enriched over the body of expressed genes and showed a positive correlation with transcript levels, although this varied and was most pronounced in the two Ectocarpales species. In contrast, H3K79me2 was broadly deposited across gene bodies but showed a mark depletion at the most highly expressed genes. In *Ectocarpus*, *S. promiscuus*, *U. pinnatifida*, elevated H3K79me2 levels were strongly associated with reduced gene expression, whereas this correlation was restricted to the TSS in the two *Desmarestia* species. While the outgroup species *S. ischiensis* showed similar hPTM deposition patterns, enrichment signals and overall correlations were weaker, with the strongest association observed for H3K9ac at the TSS and the weakest for H3K79me2. These findings indicate that hPTMs associated with active transcription function similarly as in other eukaryotic lineages, while the deposition pattern of H3K79me2 and its association with transcriptionally repressed genes differs across brown algae.

Next, we employed the Bayesian non-parametric framework hiHMM to jointly infer chromatin state maps across the six different species[21] (**Suppl. Dataset 1**). This analysis produced a probabilistic model composed of 27 emission states (**Fig. 2B, Fig. S6**), which we grouped into five broad categories based on their enrichment in specific genomic features: three TSS-associated states, five genic states, 14 mixed states distinguished by the presence or absence of H3K36me3, two silenced states, and five 'null' states lacking any of the assayed hPTMs (**Fig. 2B**, **Fig. S4B**). Closer inspection revealed both conserved and species-specific patterns of chromatin state occurrence (**Fig. S7**). For example, TSSs in all species were defined by a highly conserved and limited combination of states enriched for H3K4me3 and H3K9ac (E5, E10, E12). Gene bodies were predominantly characterised by three genic states common to all six species (E8, E2, E7). In contrast, intergenic regions and repeats exhibited more variable patterns that included a combination of mixed and silenced states. Notably, two genic states (E9, E25) were exclusively enriched in intergenic regions and repeats in *Ectocarpus* and *S. promiscuus*, suggesting a more derived chromatin landscape in the Ectocarpales lineage.

Multiple emission states can occur along the length of a single gene in various combinations, making it challenging to analyse chromatin state dynamics at the gene level. To address this, we applied our previously established approach to determine the presence of the five broad chromatin state categories at each gene across all species, resulting in 16 distinct combinations of chromatin signatures (see Materials and Methods section) (**Fig. 2C, Fig. S3C, Table S5-10**). Based on the predominant hPTMs associated with each chromatin state, we further classified the chromatin signatures into four main groups: active (S12-S16), mixed (S5-S11), repressive (S2-S4) and null (S1) (**Fig. 2C, Fig. S4C**). The relative distribution of chromatin

signatures assigned to genes was broadly conserved among species, with active signatures S15 and S16 representing the most common categories (**Fig. 2D**). We verified the relationship between the 16 chromatin signatures and gene expression in each species using paired RNA-seq data generated from the same biological material used for ChIP-seq profiling (**Fig. 2C**, **Fig. S4C**, **Fig. S8, Fig. S9**, **Table S4**). Across all brown algal species, genes assigned to active signatures consistently had higher transcript levels than those with mixed signatures, while genes associated with repressive or null signatures showed the lowest levels overall (**Fig. 2C; Fig. S9)**. In contrast, the outgroup *S. ischiensis* exhibited a much weaker correlation between chromatin signatures and gene expression, particularly for repressive signatures, suggesting a distinct regulatory system.

To better understand the relationship between H3K79me2 and transcriptional repression, we compared expression of genes marked with and without H3K79me2 in each species. In *Ectocarpus*, *S. promiscuus* and *U. pinnatifida*, H3K79me2-marked genes had significantly lower expression than unmarked genes (**Fig. 2E, Fig. S4D**). This repression was most pronounced in *U. pinnatifida*, where genes assigned chromatin signatures S12 and S14 containing H3K79me2 were largely silenced relative to their equivalent signatures S15 and S16 without H3K79me2 (**Fig. 2C**, **Fig. S4C**). In contrast, H3K79me2-marked genes in the two *Desmarestia* species did not show reduced expression compared to unmarked genes (**Fig. 2E**, **Fig. S4D**). However, metaplot analyses suggested that the association of H3K79me2 with repression may be spatially restricted to the TSS in this lineage (**Fig. S4**). Indeed, parsing genes by the presence or absence of H3K79me2 at the TSS recapitulated the repression patterns observed in the other brown algae (**Fig. 2F**, **Fig. S4E**). TSS-restricted silencing by H3K79me2 was also evident in the *U. pinnatifida* male, which also happened to have a higher proportion of null signature S1 compared to the female and other species (**Fig. 2D**). This was likely due to a high abundance of sperm cells in the *U. pinnatifida* male tissue we processed, leading us to focus the remainder of our cross-species comparisons using female samples. In the outgroup *S. ischiensis*, TSS-associated H3K79me2 also correlated with reduced gene expression, although this was limited to a relatively small subset of genes (**Fig. 2F**). These results suggest that the repressive role of H3K79me2 arose early in the brown algal lineage that has since undergone further functional diversification over the course of their evolution.

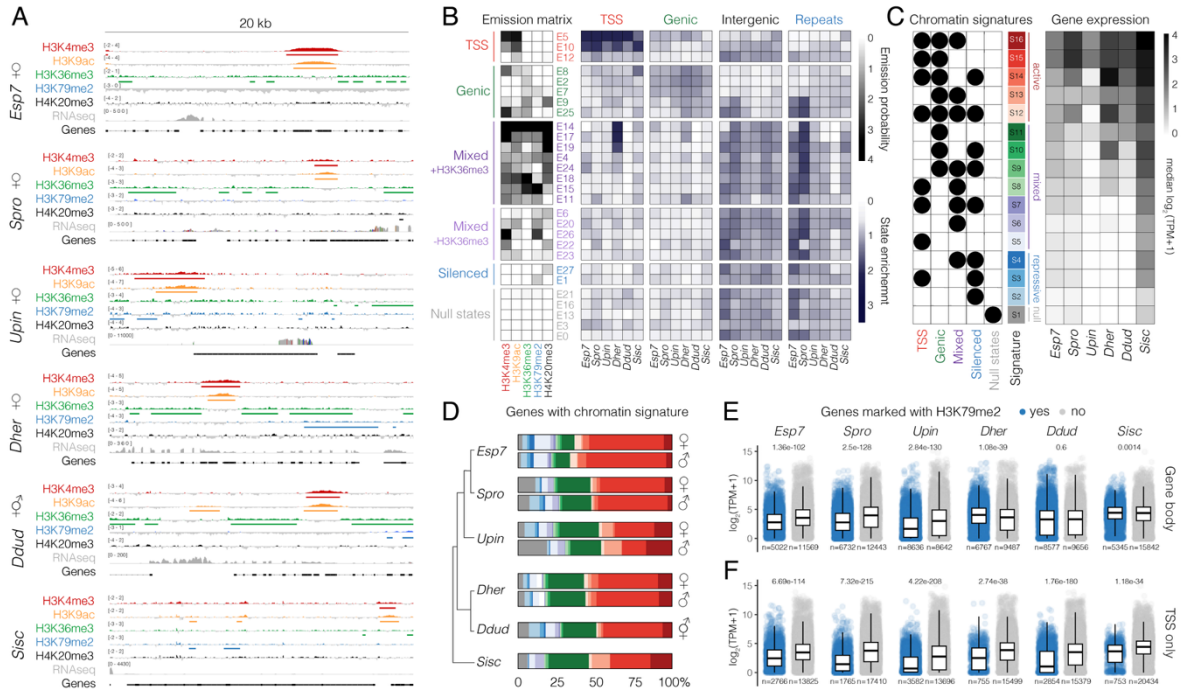



**Figure 2. The chromatin landscape across multiple brown algae**. (**A**) Representative genome browser tracks of ChIP-seq and RNA-seq datasets in each species. Data for female gametophyte is shown for dioicous species. ChIP-seq coverage is represented as the $\log_2$ ratio of IP DNA relative to histone H3, with the range indicated on each track. ChIP-seq peaks for each hPTM are indicated under their respective track (see **Fig. S4A** for male data). (**B**) A model of chromatin emission states inferred by hiHMM across all ChIP-seq datasets (left) alongside the enrichment of each chromatin state in genomic features of each species (right). Female samples are shown for dioicous species (see **Fig. S4B** for male data). (**C**) Matrix of the chromatin signatures assigned to genes based on the hiHMM emission state model (see methods) summarising the proportion of repressive, mixed, and active chromatin states within each signature. The right panel displays the median RNA-seq expression level ($\log_2$ TPM+1) of genes associated with each signature. Female samples are shown for dioicous specie s (see **Fig. S4C** for male data). (**D**) Proportion of chromatin signatures assigned to genes in each sample. (**E-F**) Gene expression with (blue) or without (grey) an H3K79me2 peak overlapping their gene body (**E**) or TSS (**F**) in each species. Female samples are shown for dioicous species (see **Fig. S4D-E** for male data). *P*-values were computed via the Kruskal–Wallis test in (**E-F**).

## Conserved function of chromatin signatures among brown algae

The brown algae are thought to have emerged around 450 million years ago, making them one of the youngest complex multicellular lineages, a fact that is reflected by the strong synteny observed across their genomes[12,22]. This prompted us to examine the evolutionary dynamics of chromatin signatures across the six species. For this, we first identified 3143 conserved single copy orthologs and assessed whether they retained similar signatures, using female data for species with separate sexes (**Table S11**)[8]. Our analysis revealed a strong overall conservation of chromatin signatures, which significantly exceeded expectation based of permutation tests assuming no conservation (**Fig. 3A**). Pairwise comparisons with model alga *Ectocarpus* further showed that the conservation of chromatin signatures correlated with phylogenetic distance (**Fig. 3B**). In *Ectocarpus*, these single copy orthologs were strongly enriched for the transcriptionally active signature S15 and largely corresponded to deeply conserved genes with

essential cellular functions and broad expression patterns, suggesting that they primarily represent housekeeping genes (**Fig. S10**). These results highlight how orthologous genes have maintained similar active chromatin organisation across brown algal evolution, which is likely driven by their constitutive expression and core housekeeping function.

We next asked whether chromatin signatures were associated with genes of similar evolutionary age and expression profile [23]. We used genomic phylostratigraphy to infer the evolutionary age of genes then assessed the distribution of their assigned signature, revealing statistically significant differences (**Fig. 3C**; **Fig. S11;** Kruskal–Wallis test; $p < 2.2e\text{-}16$ for all species). Notably, null and repressive chromatin signatures were more strongly associated with evolutionarily younger genes in *Ectocarpus* and across the other species (**Fig. 3C**; **Fig S11**). Moreover, when limiting our analysis to species-specific orphan genes, we found that these were strongly enriched in chromatin signatures with reduced expression (S1, S2, S5, S6, S11) both in *Ectocarpus* and across the clade (**Fig. 2C; Fig. 3D; Fig. S12**). This suggests that younger genes typically reside in heterochromatic regions when compared to older and more conserved genes. Differences in chromatin signatures also manifested in similar expression dynamics. Genes with null and repressive chromatin signatures exhibited more restricted expression profiles than those with active signatures, a pattern that was consistent across all species, indicating that these signatures may represent facultative heterochromatin potentially involved in developmental gene regulation (**Fig. 3E; Fig. S13**). Taken together, our findings demonstrate that chromatin signatures function similarly across 450 million years of brown algal evolution[13], revealing conserved chromatin features underlying gene regulation in this lineage.

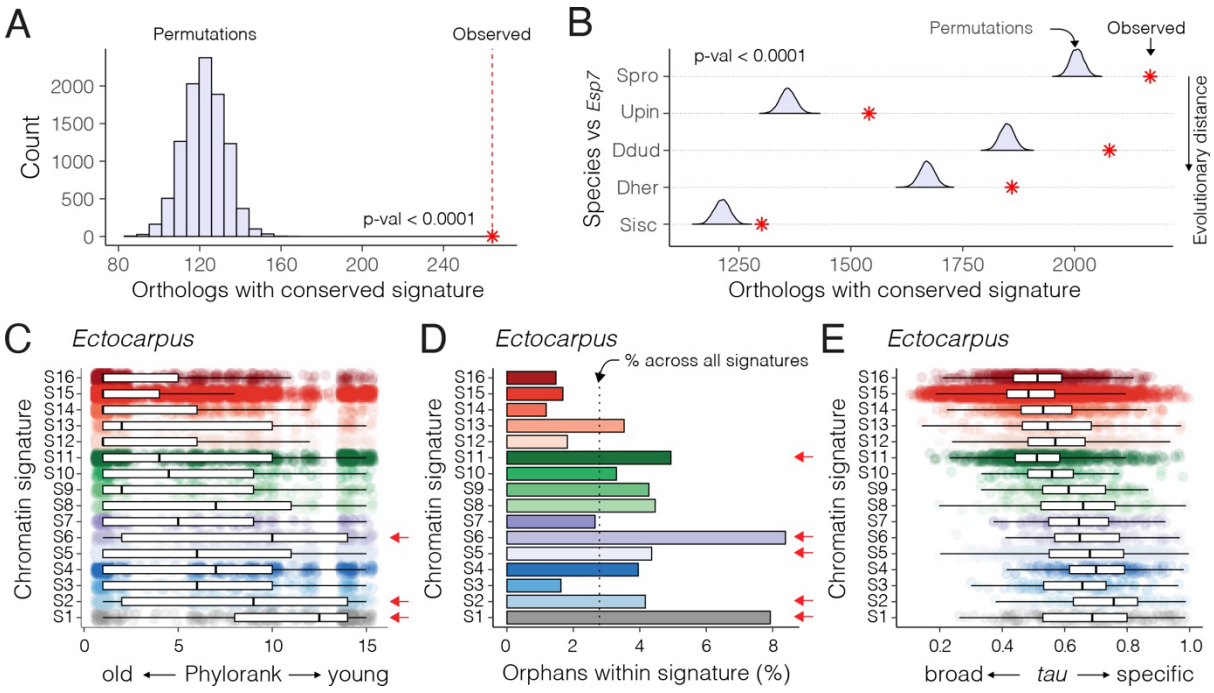



**Figure 3. Conservation and phylostratigraphy of chromatin signatures across brown algae.** (**A**) Conservation of chromatin signatures at single copy orthologs across all six species tested alongside permutations assuming no conservation. (**B**) Conservation of chromatin signatures in pairwise comparisons with the model alga *Ectocarpus* alongside permutations assuming no conservation. *P*-values were computed from the 10,000 permutation results. (**C**) Distribution of gene age across each chromatin signature in *Ectocarpus*. (**D**) Percentage

of orphan genes assigned to each chromatin signature in *Ectocarpus*. (**E**) Distribution of expression specificity scores (*tau*) for genes assigned to each chromatin signature in *Ectocarpus*.

## The chromatin landscape involved in sex determination and differentiation

UV sex chromosomes have distinctive genomic and evolutionary features due to their mode of inheritance and their characteristic sex-determining regions (SDRs) that do not recombine[22,24,25]. In *Ectocarpus*, these features are reflected by a unique chromatin landscape on the UV chromosomes, where a much higher proportion of genes are marked by repressive chromatin signatures than on autosomes[8,10]. Strikingly, this pattern was highly conserved across all five brown algal species we examined, with sex chromosome genes consistently showing reduced proportions of active signatures and increased proportions of null and repressive signatures than autosomal genes (**Fig. 4A**). The UV sex chromosomes also consistently displayed a markedly lower proportion of conserved chromatin signatures than their autosomal counterparts (**Fig. 4B**). These findings show that the UV sex chromosomes have a distinct chromatin organization across brown algae, which reflects the rapid gene turnover that is characteristic for these fast-evolving chromosomes.

We next assessed chromatin dynamics involved in sexual differentiation by comparing chromatin signatures between males and females. To capture these dynamics, we focused on major chromatin reconfigurations by comparing shifts between the four major chromatin signature groups (active, mixed, repressive and null). Across the four dioicous species, the vast majority of genes (63.2 - 87.8 %) retained the same chromatin signature group between sexes (**Fig. 4C**). Of those that were dynamic between sexes, sex-biased genes (i.e., genes that were differentially expressed between sexes) were significantly more likely to undergo major reconfiguration in *Ectocarpus* and *S. promiscuus* compared to the genomic background (**Fig. 4D**). In *D. herbacea and U. pinnatifida*, only female-biased genes were enriched for such changes (**Fig. 4D**). Genes located in the pseudoautosomal region (PAR) of the UV chromosomes were significantly enriched for dynamic chromatin reconfigurations in *Ectocarpus* and *S. promiscuus* (**Fig. 4D**). Taken together, these results suggest that major chromatin reconfigurations underlying sex-biased gene regulation are subtle and variable across brown algae and are largely restricted to specific loci, particularly female-biased genes.

Next, we focused on major chromatin reconfigurations associated with co-sexuality by comparing male and female gametophytes of the dioicous species *D. herbacea* with monoicous gametophytes of *D. dudresnayi*. Similar to the dioicous species, the vast majority of genes retained the same chromatin signature group between the co-sexual and either sex of *D. herbacea* (**Fig. 4E**). As observed during sexual differentiation in the dioicous species, major chromatin reconfigurations were preferentially associated with orthologous female-biased genes, whereas orthologous male-biased genes showed no dynamic differences in the co-sexuals compared to females, and were even less dynamic when compared to males (**Fig. 4E**). These results suggest that major chromatin reconfigurations underlying co-sexuality largely occur at female-biased genes, paralleling the situation observed in dioicy. Finally, we asked whether the transition to co-sexuality influences the conservation of chromatin signatures on the UV sex chromosomes by testing whether the reduced conservation underlying dioicy is retained when a former sex chromosome becomes an autosome (hereafter termed the 'sex-

homolog'[22]). Intriguingly, the *D. dudresnayi* sex-homolog also showed reduced conservation of chromatin signatures, indicating that this property of sex chromosomes persists after the transition to monoicy (**Fig. 4F**). Thus, the former *D. dudresnayi* sex chromosome retains molecular footprints of its past life as a sex chromosome and can leave lasting evolutionary imprints at the chromatin level.

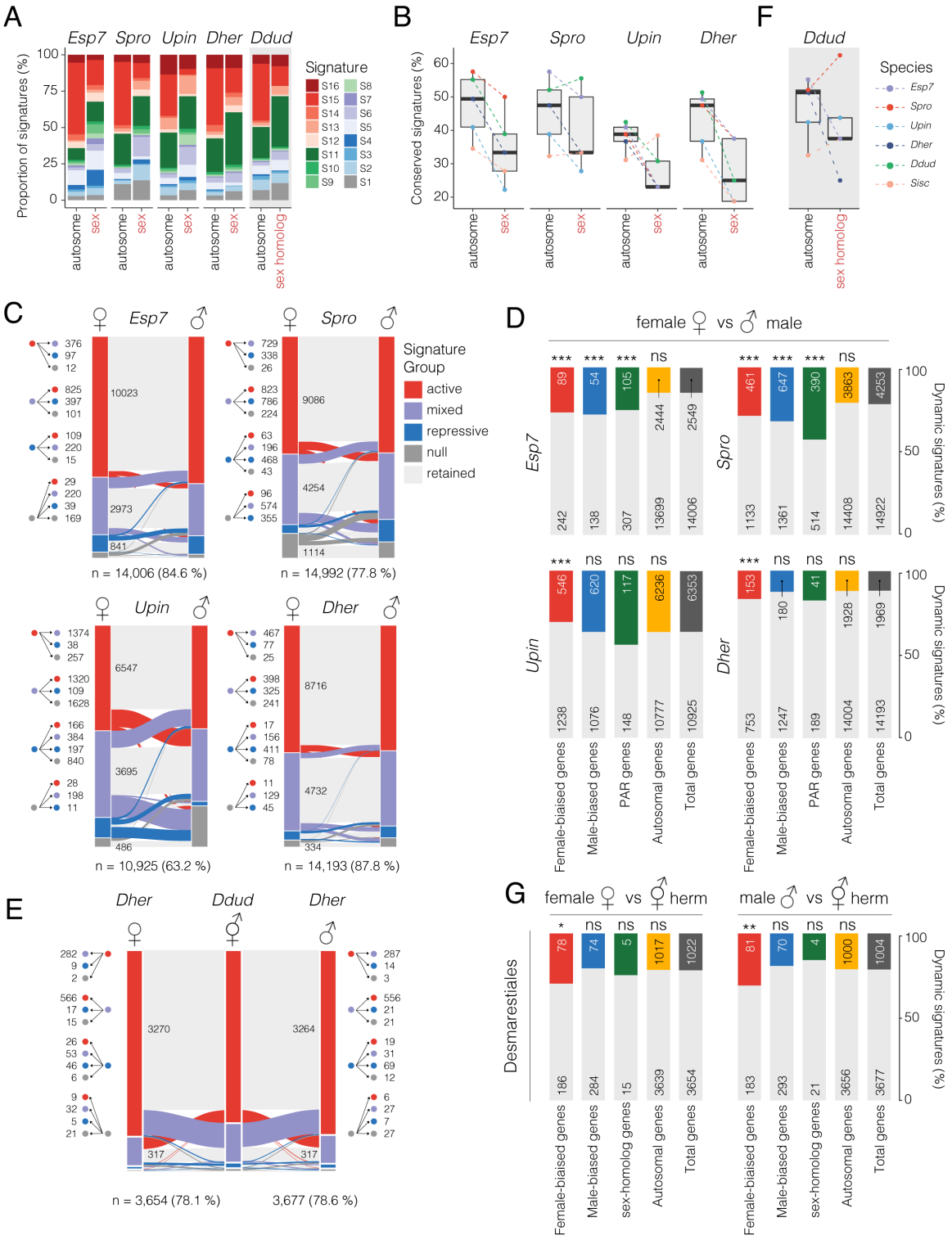



**Figure 4. Chromatin reconfiguration during sexual differentiation in brown algae with contrasting sexual systems.** (**A**) Proportion of chromatin signatures on autosomes versus sex chromosomes across dioicous brown

algal species. (**B**) Chromatin signature conservation between autosomes and sex chromosomes in each species. (**C**) Major reconfiguration in chromatin signatures between females and males. The direction of chromatin signature changes and the number of genes involved are shown on the left , with the total number of genes shown below. (**D**) Proportion of sex-biased, PAR and autosomal genes that dynamically switch chromatin signature between sexes. *P*-values of a $\chi^2$ analysis on top of each bar indicate whether the proportions observed differ significantly from the genome average. (**E**) Major reconfiguration in chromatin signatures between females and males versus co-sexual *Desmarestia* species. The direction of chromatin signature changes and the number of genes involved are shown on the left, with the total number of genes shown below. (**F**) Chromatin signature conservation across all autosomes versus the sex homolog (ancestral sex chromosome) in the co-sexual species *Desmarestia dudresnayi*. (**G**) Proportion of sex-biased, PAR and autosomal genes that dynamically switch chromatin signatures in *D. herbacea* (dioicous) versus *D. dudresnayi* (monoicous).

## Epigenetic control in an ancestor of the brown algae

The outgroup *S. ischiensis* stood out among the six species we analysed due to its weaker correlation between chromatin signatures and gene expression, and its relatively low number of H3K79me2-marked genes. This prompted us to explore other potential regulatory mechanisms that may operate in *S. ischiensis*. Our phylogenetic analysis revealed that, unlike brown algae, *S. ischiensis* encodes a MET1 ortholog (**Fig. 1A**), leading us to speculate that DNA methylation may regulate gene expression in this species. To explore this, we called different forms of DNA methylation from Oxford Nanopore long-read sequences used to assemble the *S. ischiensis* genome. The level of 4-methylcytosine and 6-methyladenine calls are unlikely to be genuine in *S. ischiensis* since their genome-wide levels were very low (< 5%) and comparable to those detected on the plastid genome (**Fig. S14**). In contrast, 5-methylcytosine in a CG context (5mCG) was highly abundant, reaching almost 80% genome-wide and far exceeding background levels on the plastid genome (**Fig. 5A-B**). By comparison, 5mC in CHG and CHH contexts was much lower and showed no preference for TEs as they do in plants, suggesting that these methylation patterns are unlikely to have a specific function in *S. ischiensis* (**Fig. S14**).

Closer inspection of 5mCG in *S. ischiensis* revealed discrete regions with a complete loss of methylation around the promoter of protein-coding genes (**Fig. 5A**), which is reminiscent of unmethylated CpG "islands" found in mammalian genomes[26,27]. These demethylated islands coincided with the enrichment of TSS-associated H3K4me3 and H3K9ac, with the upstream region of genes consistently having reduced 5mCG levels relative to other genomic regions (**Fig. 5A-B**). Hierarchical clustering of 5mCG levels alongside the six profiled hPTMs revealed two main clusters distinguished by demethylated promoter regions and the concurrent enrichment of H3K4me3 and H3K9ac (**Fig. 5C**). Genes in this "active" cluster had higher transcript levels compared to genes without promoter demethylation, consistent with the deposition of active hPTMs (**Fig. 5D**). These results explain the weaker correlation of chromatin signatures in *S. ischiensis* (see **Fig. 2C**) and suggest that active DNA demethylation at promoters may have played a key regulatory role prior to the loss of DNA methylation during early brown algal evolution.

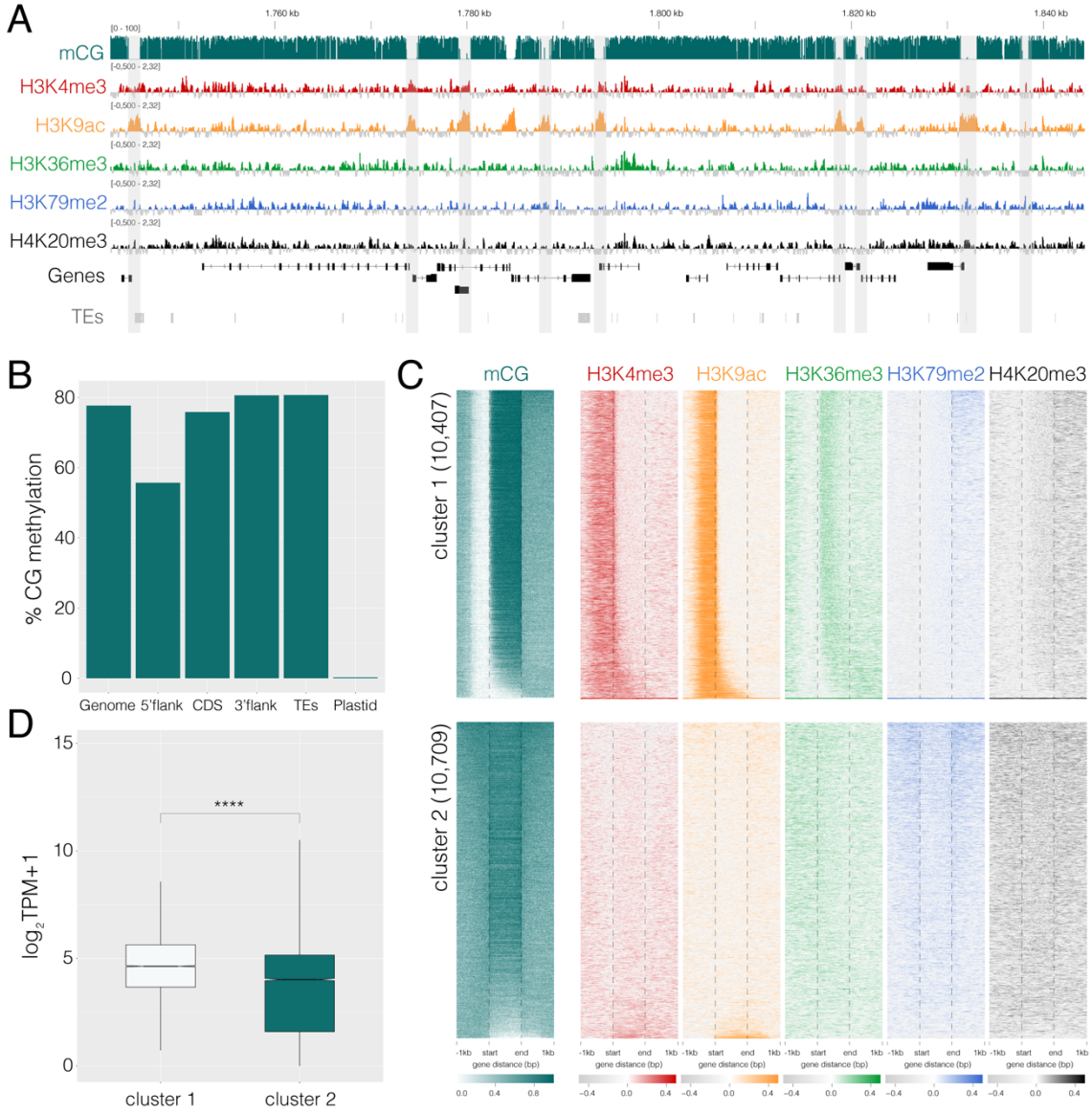



**Figure 5**. **Epigenetic landscape in the closest outgroup of the brown algae.** (**A**) Genome browser view of ChIP-seq and CG methylation tracks in *Schizocladia ischiensis*. Demethylated CpG-like islands are shown with grey shading. (**B**) Average genome-wide CG methylation levels at different genomic features of the nuclear genome and plastid genome. (**C**) Chromatin and CG methylation landscape over *S. ischiensis* genes clusters based on the presence (cluster 1) and absence (cluster 2) of CG methylation at promoter regions. (**D**) Gene expression level of genes forming part of cluster 1 and cluster 2 shown in panel C.

## Discussion

Brown algae are the most recent lineage to have independently evolved complex multicellularity, making them an important case study to understand how regulatory mechanisms evolve during the emergence of organismal complexity. Our comparative analysis across the Phaeophyceae and their closest relatives reveals that this transition was accompanied by a fundamental shift in chromatin regulation, where canonical repressive pathways involving PRC2, H3K9 and DNA cytosine methylation are largely dispensable across the clade. Although

the presence of DNA methylation is highly variable across eukaryotes[28], H3K27 methylation is largely ubiquitous and essential in animals, fungi, land plants, and even closely related Stramenopiles[29,30]. In brown algae, the concurrent loss of DNA, H3K9 and H3K27 methylation is unprecedented and underscores the unique evolutionary trajectory of this lineage.

Our analysis of the outgroup *S. ischiensis* provides an important perspective on the ancestral regulatory landscape of the brown algae. We show that the presence of DNA methyltransferases is accompanied by high genome-wide levels of 5mCG methylation. Interestingly, we reveal a strong link between promoter demethylation and the accumulation of active histone marks at expressed genes, suggesting a regulatory model in which DNA methylation plays a central role in gene repression. The localised demethylation of promoter regions and the coordinated enrichment of active histone marks point to a regulatory system that is strikingly similar to vertebrates rather than to modern-day brown algae[26,27]. The absence of PRC2 orthologs in *S. ischiensis* suggests that H3K27me3-mediated regulation is likely to have been lost early in brown algal evolution. The eventual loss of both DNA methylation and PRC2 in the common ancestor of brown algae, which is exceptionally rare among multicellular eukaryotes, would have necessitated the emergence of alternative repressive systems, providing fertile ground for the adaptation of DOT1-mediated H3K79 methylation into a key repressive pathway.

The expansion and diversification of DOT1-like enzymes into lineage-specific families is a distinguishing feature of Phaeophyceae evolution. In yeast and animals, DOT1-dependent H3K79 methylation regulates diverse processes including gene expression, replication initiation, DNA damage response, microtubule reorganization and chromosome segregation[17]. DOT1 enzymes also promote heterochromatin formation by regulating pericentromeric transcription of satellite repeats, where bursts of transcription are required to establish and maintain long-term silencing[31,32]. We speculate that this role in heterochromatin formation could have been co-opted during brown algal evolution to give rise to its repressive role in the modern-day. H3K79 methylation has similarly been implicated in gene repression in other eukaryotic lineages[7], underscoring a capacity for the DOT1 pathway to be independently recruited to regulate silencing during eukaryotic evolution.

By investigating chromatin landscapes across diverse brown algal species, we established a defined set of combinatorial chromatin states (signatures) that are highly predictive of gene expression across the clade. Active transcription was consistently associated with H3K4me3, H3K9ac, and H3K36me3, whereas H3K79me2 and H4K20me3 were associated with reduced transcription. The localisation of H3K79me2 and its perceived repressive role varied among the brown algal clades, further highlighting the dynamic nature of this pathway. In addition to transcriptional states, chromatin signatures were also strongly correlated with the evolutionary age of genes and the breadth of their expression. Notably, repressive chromatin signatures were highly enriched among young species-specific genes with restricted gene expression patterns, a pattern mirrored across a wide range of eukaryotes[33–35]. This supports the idea that heterochromatic regions may serve as a cradle for the emergence of novel genes[22], with silencing followed by eventual reprogramming and expression in reproductive cell types providing a route for the evolution and selection of new gene functions[35–37].

Our findings have also shed light on the regulation of sex determination and sex chromosome evolution in brown algae. Brown algal UV sex chromosomes consistently show enrichment of repressive chromatin and reduced conservation of chromatin signatures compared to autosomes. These features are consistent with suppressed recombination and gene turnover in the sex-determining regions[15,38–40] which appear to leave a lasting epigenomic footprint. Interestingly, even after the transition from dioicy to co-sexuality in the Desmarestiales, the sex-homolog retains a distinct chromatin profile, suggesting that the epigenomic legacy of sex chromosome identity persists long after recombination resumes. Chromatin modifications thus not only reflect current transcriptional states but may also shape the long-term evolutionary trajectory of the sex-homolog during transitions towards co-sexuality by constraining or biasing subsequent regulatory evolution. This echoes findings from plants and animals where epigenetic silencing marks, dosage-compensation mechanisms and heterochromatin expansion continues to influence genome function long after sex chromosome turnovers or fusions[41–44]. Such epigenomic legacies may therefore represent a general principle in the evolution of sex determination systems. We also uncovered major chromatin reconfigurations associated with sexual differentiation across brown algae. *Ectocarpus* and *S. promiscuus*, both species with limited sexual dimorphism, showed the most pronounced localised chromatin reconfigurations at PAR genes and sex-biased genes. In other species, such reconfigurations consistently occurred at female-biased genes, highlighting augmentation of the chromatin reconfiguration on the female program as the principal driver of sexual differentiation in brown algae with stronger sexual dimorphism, including those with co-sexual systems.

In summary, our findings highlight evolutionary innovations in the chromatin toolkit that accompanied the emergence of complex multicellularity in brown algae, where the loss of canonical repression pathways and the rise of DOT1/H3K79 methylation established a novel regulatory system that now underpin development and reproduction in this vital and unique eukaryotic lineage.

## Methods

### Genome screening and orthology inference

We initially performed genome screening to identify the components of epigenetic regulation in brown algae using blastp[45] (default parameters) by leveraging recently published genomic data of species in the brown algal lineage (Phaeophyceae) and early-diverging Ochrophytina[12,15] along with representatives from across major eukaryotic groups. We selected five species covering the brown algal phylogenetic diversity and one outgroup species. Orthology inference was performed on these six species using OrthoFinder v2.5.5[46].

### Biological material

Gametophytes of the five brown algal species and the outgroup *S. ischiensis* were cultivated in 90cm Petri dishes (Corning® Gosselin™ BH90B-102) containing at least 10 individuals, with Provasoli enriched seawater as described in [19,47]. Fertile individuals were harvested with a 70µm strainer, then rinsed with seawater and dry with a paper towel for further processing. Light and temperature conditions were optimised for fertility, as described in **Table S1**.

## hPTM profiling

**Histones extraction**. Histones were extracted from 0.5g of frozen algae, pulverized in liquid nitrogen. The powder was then homogenised in 40mL of M1 buffer (10mM Na Phosphate Buffer pH7, 100mM NaCl, 1000mM Hexylene Glycol, 10mM b-mercaptoethanol, 1X cOmplete Proteinase inhibitor Cocktail (Roche). After filtering through 2 layers of Miracloth (Milipore #475855), each sample was centrifugated at 2000 × g for 10 minutes at 4 °C. The pellet was carefully resuspended in 80mL M2 buffer (10mM Na Phosphate Buffer pH7, 100mM NaCl, 10mM MgCl2, 1000mM Hexylene Glycol, 0.1% Triton X-100, 10mM b-mercaptoethanol, 1X cOmplete™, EDTA-free Protease Inhibitor Cocktail (Roche, #CO-RO) twice. This pellet was then incubated with Extraction buffer (1000mM CaCl2, 20mM Tris-HCl pH7.25, 1x cOmplete cOmplete™, EDTA-free Protease Inhibitor Cocktail) for 10 minutes on ice. 0.3N of 37% HCl was added thereafter followed by centrifugation at 10 000 × g for 5 minutes at 4 °C. The resulting supernatant was collected in a fresh Protein LoBind® Tubes Eppendorf tubes. Histones were precipitated with 20% tricholoacetic acid (TCA) and incubated for 10 minutes before 13 000 × g centrifugation at 4 °C for 30 minutes, followed by successive washes with 20% ice-cold TCA, ice-cold acetone supplemented with 0.2% HCl and ice-cold acetone. The pellet was dried at room temperature and resuspended in miliQ water overnight at 4 °C.

**Western blot**. Histone samples were supplemented with Laemmli 2X and 100mM of DTT and NaOC until blue coloration was observed and incubated at 95 °C for 5 minutes. Histone PTMs were detected on a 15% handcast SDS-PAGE gel, using the same antibodies listed belowy as in the ChIP experiment. For H3, 3–15 µg of tissue-equivalent sample was loaded onto the gel. Histone samples, corresponding to 3–15 µg of tissue equivalent for H3, were loaded. For H3K9me3, around 10 ug of tissue-equivalent sample twice that amount was used, and for other histone marks, approximately 40 ug of tissue-equivalent sample four times or more. Proteins were transferred onto a 0.45µm nitrocellulose membrane (0.45µm, BioRad, #1620113) on a Trans-Blot® Turbo™ Transfer System (BioRad, #1704150). Membranes were blocked in 5% milk in 1× PBS-T for 30 minutes. Primary antibodies were diluted in 5% milk in 1× PBS-T and incubated for 1 hour at room temperature. These included rabbit anti-H3 (Histone H3 (D2B12) XP Rabbit mAb (ChIP Formulated), Cell Signaling Technology #4620S), anti-H3K4me3 (Tri-Methyl-Histone H3 (Lys4) (C42D8) Rabbit mAb, CST #9751S) , anti-H3K9ac (Acetyl-Histone H3 (Lys9) (C5B11) Rabbit mAb, CST, #9649S), anti-H3K79me2 (Di-Methyl-Histone H3 (Lys79) (D15E8) XP Rabbit mAb, CST, #5427S), and anti-H4K20me3 (Tri-Methyl-Histone H4 (Lys20) (D84D2) Rabbit mAb, CST, #5737S) at 1:1000 dilution, and anti-H3K36me3 (AbcamTri-Methyl-Histone H3 (Lys36) (D5A7) XP Rabbit mAb, Abcam, #4909S) at 1 µg/mL. Membranes were incubated with HRP-conjugated anti-rabbit secondary antibody (1:2000, CST #7074S). After further washes, membranes were developed using a 1:1 mix of Trans-Blot® Turbo™ Transfer System (BioRad, #1704150). Images were captured with ChemiDoc Imaging System from BioRad.

**Mass spectrometry**. Samples for liquid chromatography-tandem mass spectrometry (LC-MS/MS) were prepared by migrating the extracted histone on a 14% SDS-polyacryamide gel at 100V for 10 minutes. The gel was dyed with LabSafe GEL Blue™ (G-BIOSCIENCES,

#786-35) following the manufacturer instruction. This was followed by LC-MS/MS, performed by coupling a Vanquish Neo LC system (Thermo Scientific) to an Orbitrap Astral mass spectrometer, interfaced by a Nanospray Flex ion source (Thermo Scientific). In a subsequent round of analyses, a RSLCnano system (Ultimate 3000, Thermo Scientific) to an Orbitrap Exploris 480 mass spectrometer (Thermo Scientific) was additionnally employed.

On the Vanquish Neo LC system, peptides were injected onto a C18 column (75 µm inner diameter x 50 cm double nanoViper PepMap Neo, 2µm, 100Å, Thermo Scientific) regulated also at 50 °C, and separated with a linear gradient from 100% buffer A' to 28% buffer B at a flow rate of 300 nL/min over 104 minutes. The Orbitrap Astral mass spectrometer was run in Data Dependent Acquisition (DDA) mode and MS full scans were performed in the ultrahigh-field Orbitrap mass analyzer in ranges m/z 380–1200 (resolution of 240 000 at m/z 200; maximum injection time 100 ms; AGC 300%). For the Astral MS/MS spectra, the top N most intense ions were isolated and subjected to further fragmentation via high energy collision dissociation (HCD) activation with the auto gain control (AGC) target set to 100%. We selected ions with charge state from 2+ to 6+ for screening. Normalised collision energy (NCE) was set at 30 and the dynamic exclusion at 20s.

On the RSLCnano system, peptides were trapped on a C18 column (75 µm inner diameter × 2 cm; nanoViper Acclaim PepMapTM 100, Thermo Scientific) with buffer A (2/98 CH3CN/H2O in 0.1% formic acid) at a flow rate of 2.5 µL/min over 4 minutes. Separation was performed on a 50 cm × 75 µm C18 column (nanoViper Acclaim PepMapTM RSLC, 2 µm, 100Å, Thermo Scientific) regulated to a temperature of 50 °C with a linear gradient of 2% to 30% buffer B (100% CH3CN in 0.1% formic acid) at a flow rate of 300 nL/min over 91 minutes. On the Orbitrap Exploris 480 mass spectrometer, full scans were performed in ranges m/z 375–1500 (resolution of 120 000 at m/z 200; maximum injection time 25 ms; AGC 300%) and the top 20 most intense ions were isolated and subjected to further fragmentation via HCD activation at resolution of 15 000 with the AGC target set also to 100%. We also selected ions with charge state from 2+ to 6+. NCE was set at 30 and with a dynamic exclusion of 10s.

The resulting LC-MS/MS data was searched against the species-specific histone sequences using Mascot[48]. Enzyme specificity was set to trypsin and a maximum of five-missed cleavage sites were allowed. Oxidized methionine, carbamidomethylated cysteine, N-terminal acetylation, acetylation, methylation (mono, di and tri) of lysine, methylation (mono and di) of arginine, methylation of glutamic acid and aspartic acid were set as variable modifications and with a maximum of nine modifications for all Mascot searches. Maximum allowed mass deviation was set to 10 ppm for monoisotopic precursor ions and 0.02 Da for MS/MS peaks. The resulting Mascot files were further processed using myProMS (v.3.10; https://github.com/bioinfo-pf-curie/myproms)[49].

## ChIP-seq

To map hPTMs to the genome, we performed ChIP-seq to detect the enrichment of H3K4me3, H3K9ac, H3K36me3, H3K79me2, and H4K20me3. Each sample was prepared from approximately 0.6 g of semi-dry algal tissue (~600 individuals), which was then fixed in seawater containing 1% freshly prepared formaldehyde for 10 minutes. The fixed sample was

quenched with 400 mM glycine in 1× PBS, followed by rinsing with fresh seawater to remove residual formaldehyde. Nuclei were isolated by grinding the cross-linked tissue in liquid nitrogen and resuspending the powder in a nuclear isolation buffer containing 0.1% Triton X-100, 125 mM sorbitol, 20 mM potassium citrate, 30 mM $MgCl_2$, 5 mM EDTA, 5 mM β-mercaptoethanol, 55 mM HEPES (pH 7.5), and 1× EDTA-free protease inhibitor cocktail (Roche #CO-RO). The suspension was homogenized using a Tenbroeck Potter, filtered through Miracloth (Millipore #475855), and centrifuged at 3000 × g for 10 minutes at 4 °C. The nuclear pellet was washed twice with the same buffer and once more with buffer lacking Triton X-100, conserving the centrifuge parameters. Nuclear pellets were then lysed in 1ml of nuclear lysis buffer total (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8, and protease inhibitors). Chromatin was fragmented via sonication using a Covaris E220 Evolution sonicator (settings: 25% duty cycle, 75 peak power, 200 cycles/burst, 900 s duration at 4 °C) in 8 microTUBE AFA Fiber Snap-Cap tubes. Cellular debris was cleared by centrifugation at 14 000 × g for 5 minutes at 4 °C. The resulting chromatin-containing supernatant was diluted 1:10 with ChIP dilution buffer (1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8, 167 mM NaCl, and 1× EDTA-free protease inhibitor cocktail (Roche #CO-RO)). Diluted chromatin was distributed into DNA LoBind tubes (Eppendorf) and incubated overnight at 4 °C with a 1:500(v/v) antibody on a rotator set at 10 rpm. Antibodies were sourced from Cell Signaling Technology (anti-H3: #4620, H3K4me3: #9751S, H3K9ac: #9649S, H3K79me2: D15E8, H4K20me3: #5737S) and Abcam (H3K36me3: ab9050). Immunoprecipitation was carried out using a 1:1 mixture of protein A and G Dynabeads (Thermo Fisher Scientific #10004D and #10002D). Following binding and sequential wash steps, immune complexes were eluted in 100 µl of Direct Elution Buffer (0.5% SDS, 5 mM EDTA, 10 mM Tris-HCl pH 8, 300 mM NaCl). Cross-link reversal was achieved by incubating samples at 65 °C overnight with intermittent shaking. DNA was purified following digestion with Proteinase K (Fisher Scientific #11826724) and RNase A (Roche #10109142001). DNA extraction was performed using phenol/chloroform/isoamyl alcohol (25:24:1), followed by centrifugation at 13 800 × g for 15 minutes at 4 °C. The aqueous phase was transferred to fresh DNA low binding tubes, mixed with 1.25 ml of 100% ethanol, 50 µl of 3 M sodium acetate (pH 5.2), and 4 µl of glycogen (20 mg/ml), and incubated at –80 °C for at least 1 hour (or overnight) for DNA precipitation. DNA was pelleted by centrifugation at 13 800 × g for 15 minutes at 4 °C, washed with 70% ethanol, and centrifuged again under the same conditions. Pellets were air-dried and resuspended in 0.1× TE buffer. Library preparation was conducted using the NEBNext® Ultra™ II DNA Library Prep Kit (New England Biolabs #E7645S), and sequencing was carried out on the Illumina HiSeq 3000 platform, targeting 20 millions of 150-bp paired-end reads per sample.

To process the ChIP-seq data, we used nf-core/chipseq v2.0.0 with default options [50]. Publicly available datasets from wild-type male and female gametophytes of *Ectocarpus*[8] were retrieved and processed using the same workflow for consistency. Biological replicates for each species were aligned to their corresponding reference genome (see **Table S1**). Replicates showing high correlation, as determined by Spearman's coefficient using `multiBamSummary` and `plotCorrelation` from deepTools v3.5.1[51] were merged using samtools merge for downstream analyses. Peaks were called with `macs2 (defaul parameters)`.

Normalised $\log_2$ coverage tracks, relative to total H3, were generated using `deepTools bamCompare` with a bin size of 10 bp, `--scaleFactorsMethod readCount`, and `--operation log2`. Using the deepTools, we computed the correlation matrices via `multiBigwigSummary` (in bins mode) followed by `plotCorrelation`. Genome-wide signal profiles were visualized in IGV v2.18.4.

## RNA-seq

RNA-seq data was generated from culture with same conditions to match the histone PTM data with the gene expression data. Each RNA-seq were carried out in triplicates. For each replicate, approximately 10 mg of algal tissue was gently blotted dry and immediately flash-frozen in liquid nitrogen. Total RNA was extracted following the method described by[10]. Briefly, frozen tissue was ground in liquid nitrogen and incubated at 65 °C in 700 µL of preheated CTAB3 extraction buffer (100 mM Tris-HCl pH 8, 1.4 M NaCl, 20 mM EDTA pH 8, 2% CTAB, 2% PVP, and 1% β-mercaptoethanol). The lysate was vortexed and maintained at 65 °C for 5–20 minutes. Phase separation was achieved by extraction with chloroform:isoamyl alcohol (24:1), followed by two rounds of centrifugation at 10 000 × g for 15 minutes at 4 °C. RNA was precipitated overnight at –20 °C using 3 M LiCl and 1% β-mercaptoethanol, pelleted by centrifugation at 10 000 × g for 1 hour at 4 °C, washed with cold 70% ethanol, and resuspended in RNase-free water. Genomic DNA contamination was removed using the TURBO DNase Kit (Thermo Fisher, AM1907). RNA-seq libraries were prepared using the NEBNext® Ultra™ II Directional RNA Library Prep Kit (New England Biolabs, E7760S) and sequenced on the Illumina Next Seq 2000 platform, generating 25–30 million 150-bp paired-end reads per sample.

Reads were processed using the nf-core/rnaseq pipeline v3.12.0 [50]. Genome assembly versions are described in **Table S1**. Publicly available RNA-seq datasets from *Ectocarpus sp.7* male and female gametophytes[8] were reprocessed using the same workflow for consistency. Reproducibility across biological replicates was confirmed via Spearman correlation. Differential expression analysis between male and female was performed with DESeq2 (v1.42.1)[52], identifying differentially expressed genes (DEGs) based on a $|\log_2 \text{fold change}| \geq 1$ and an adjusted p-value $< 0.05$.

## Generation of histone mark profiles

To investigate the distribution of histone modifications across genes grouped by expression level, we used a custom bash script to automated signal processing and plotting using with deepTools v3.5.1 within a conda environment. Genes were divided into five groups: one group comprising genes with zero expression (0 TPM), and the remaining genes divided into four groups according to expression level quartiles one where there were no expression and the rest split by quartiles (calculated for each sample). Gene coordinates grouped by expression quantiles were provided as BED files, and ChIP-seq signal was taken from precomputed bigWig tracks. Signal matrices were generated using `computeMatrix scale-regions -a 1000 -b 1000 -bs 100 –skipZeros`, which summarized ChIP-seq signal over gene bodies (represented as 5 kb) and 1 kb flanking regions, with a bin size of 100 bp. Regions with zero signal were excluded. The resulting matrices were visualized using `plotProfile`

`--plotType se`, producing average signal profiles per quantile group.

## Chromatin emission state and signature inference

**Detection of emission states with hiHMM**. The five chromatin marks were analysed using hiHMM [21] to annotate each genome with emission states. Input files for hiHMM were generated in several steps following the recommendations from [21]. First, BedGraph files were produced using `bamCompare -bs 200 --scaleFactorsMethod readCount --pseudocount 0.5 --operation log2 -of "bedgraph"`. Then BedGraph files were normalized prior to modeling using a short Python script that applied `sklearn.preprocessing.StandardScaler.fit_transform()` to the signal column (the 4th). Quality check involving visual inspection of quantile-quantile plots confirmed standardization. Normalized BedGraphs were reformatted into chromosome-wise matrices with genomic bins as rows and samples as columns and file and chromosome names were standardized as required.

The hiHMM model was trained on a reduced subset of chromosomes (one autosome per sex, plus one male and one female sex chromosome or sex homolog when available), which reduces computational complexity without affecting model quality according to the authors[21]. For each species, we selected the two longest scaffolds together with the available sex chromosome(s). The list of training chromosomes and run parameters are provided in **Table S11**. The optimised model initiates with $K_0 = 7$ and ends up with K = 27 states plus one "E0" state containing regions where no reads were assigned. The decoding part was carried on all scaffolds. The decoding part was carried on all scaffolds. The emission matrix can be found in **Fig. 2B** and transition matrix can be found in **Fig. S13** as well as the input files used in **Supplemental Dataset 1**. Model 2 was favored, as Model 1 did not provide an integrated or streamlined representation of the data. To examine the hiHMM model, we adapted the advanced functions of ChromHMM[53]. The optimization of hiHMM model was obtained using `ChromHMM CompareModels` function where hiHMM model was tested for different value of $K_0$. From these comparisons, we decided to use `K0 = 7` (as the default). The overlap enrichment of genomic features was produced with `ChromHMM OverlapEnrichment`.

**Definition of chromatin signatures**. To streamline the analysis, emission states assigned to each gene were consolidated into five broader categories based on their predominant emission enrichment with genomic features as done in[54]. States showing strong enrichment at transcription start sites (TSS) were classified under the "TSS" category (E5, E10, E12). Emission states predominantly associated with gene bodies were grouped into the "Genic" category (E8, E2, E7, E9, E25). States displaying mixed enrichment patterns—with or without H3K36me3—were collectively grouped under "mixed," due to the absence of clear feature-specific enrichment (E14, E17, E19, E4, E24, E18, E15, E11, E6, E20, E26, E22, E23). A distinct "Silenced" category was defined for states marked exclusively by H3K79me2 and H4K20me3 (specifically E27 and E1). Finally, states characterized by minimal or no detectable histone modification signals were assigned to the "Null states" group (E0, E3, E13, E16, E21). Each gene was then annotated based on its combination of these five chromatin categories, resulting in a defined set of unique chromatin profiles referred to as 'chromatin signatures'.

Genes were assigned to the null signature group (S1) if they were associated solely with low-signal states (E0, E3, E13, E16, E21), and no other chromatin category. In total, this classification yielded 16 distinct chromatin signatures (S1–S16) across the six analysed genomes. Signatures were then labeled following the increasing gene expression median across all species.

Chromatin signatures of genes were extracted from **Table S5-10** and compared between samples. Class transitions were summarized with sankey plots using `ggalluvial::geom_alluvium()` to visualise flux as well as stacked bar plot using `ggplot2::geom_col()` in R. Enrichment of chromatin changes was examined by different category type such as gene location (autosomes, PAR, SDR, sex-homolog chromosome) and expression bias (female-biased, male-biased, unbiased). Statistical relevance was assessed with chi-square test using `stats::chisq.test()` in R, comparing the count of one category against the total genes. P-values were Bonferroni adjusted using `p.adjust(method = "bonferroni")`.

## Evolutionary analysis of chromatin signatures

We analysed the overall conservation of chromatin signatures by counting the number of one-to-one orthologs with the same chromatin signature label across all six species. To examine the significance of the observed value, we performed a permutation test by reshuffling the chromatin signature labels across orthologs for each species and recomputing the number of overlaps. The permuted values represent the null assumption of no conservation of chromatin signatures. The same procedure was followed for pairwise comparisons (against *Ectocarpus*). These analyses were enabled by rsample v1.2.1 (https://rsample.tidymodels.org). Female chromatin signature data was used for dioicous species in this and following analyses. *P*-values were computed from the 10,000 permutation results by comparing the observed statistic to the empirical null distribution.

To examine the influence of sex chromosomes on the signature conservation, we calculated the percentage of observed cross-species pairwise overlaps separately for autosomes and sex chromosomes for each focal species. We only considered the chromosome type in the focal species and not the target species of the comparison due to the high evolutionary turnover rate of one-to-one orthologues in the sex chromosome. This analysis included the former sex chromosome of *D. dudresnayi* but not *S. ischiensis* since the sexual system is unknown. Statistical comparisons between autosomes and sex chromosomes were conducted using paired, one-tailed Wilcoxon rank sum tests.

To profile the gene-wise evolutionary signal associated with each chromatin signature, we performed genomic phylostratigraphy using GenEra[23] to infer the evolutionary age of each gene, with the resulting phylostratigraphic rank (phylorank) ranging from 1 (cellular organisms) to 15 (species-level) in *Ectocarpus*. It should be noted that the phylorank of the youngest genes differ between species due to differences in the presence of genomic data at each taxonomic node. For the orphan gene analysis, gene models with the highest phylorank for each species were considered as orphan genes, which we then obtained their prevalence (in %) in each chromatin signature.

Alongside the evolutionary signals, we profiled the expression variability associated with each chromatin signature. Using a recently published developmental RNA-seq dataset of *Ectocarpus*[55] we calculated *tau* scores[56] to estimate the expression specificity of genes, which ranges from 0 (broadly-expressed) to 1 (developmental stage-specific expression), using the median expression values for each gene across replicates. For other species, which lack comparably comprehensive RNA-seq datasets, we calculated expression variability using the RNA-seq data produced in this study as input. Specifically, we computed the relative entropy (using `KL.empirical()` from entropy v1.3.2) for each gene between the observed $\log_2$(TPM+1) expression values across replicates and a null assumption of uniform expression for each gene. To reduce the effect of noise in lowly expressed genes, genes with mean TPM below 3 were discarded for this analysis.

## Data analytics and graphics

Statistical analysis, data processing and visualisation were done using R v4.3.3 with the following packages: tidyverse v2.0.0[57] (ggplot2 v3.5.2, dplyr v1.1.4, tidyr v1.3.1, tibble v3.3.0, stringr v1.5.1, readxl v1.4.3), rtracklayer v1.64.0[58], duckplyr v0.4.1 (https://duckplyr.tidyverse.org/), GenomicRanges v1.56.2, ggalluvial v0.12.5 (http://corybrunson.github.io/ggalluvial/), ggrastr v1.0.2 (https://github.com/VPetukhov/ggrastr), and rsample v1.2.1 (https://rsample.tidymodels.org). Statistical tests and significance levels are indicated in the text, figure legends and **Table S12**. Scripts are provided at https://github.com/jerovign/PhaeoChromo.

## Data availability

Mass spectrometry data are available via ProteomeXchange under identifier PXD065559. ChIP-seq and RNA-seq short-read data has been uploaded to NCBI under BioProject PRJNA1328953. Supplemental Dataset 1 can be accessed at https://doi.org/10.17617/3.TDGYHS.

## Contributions

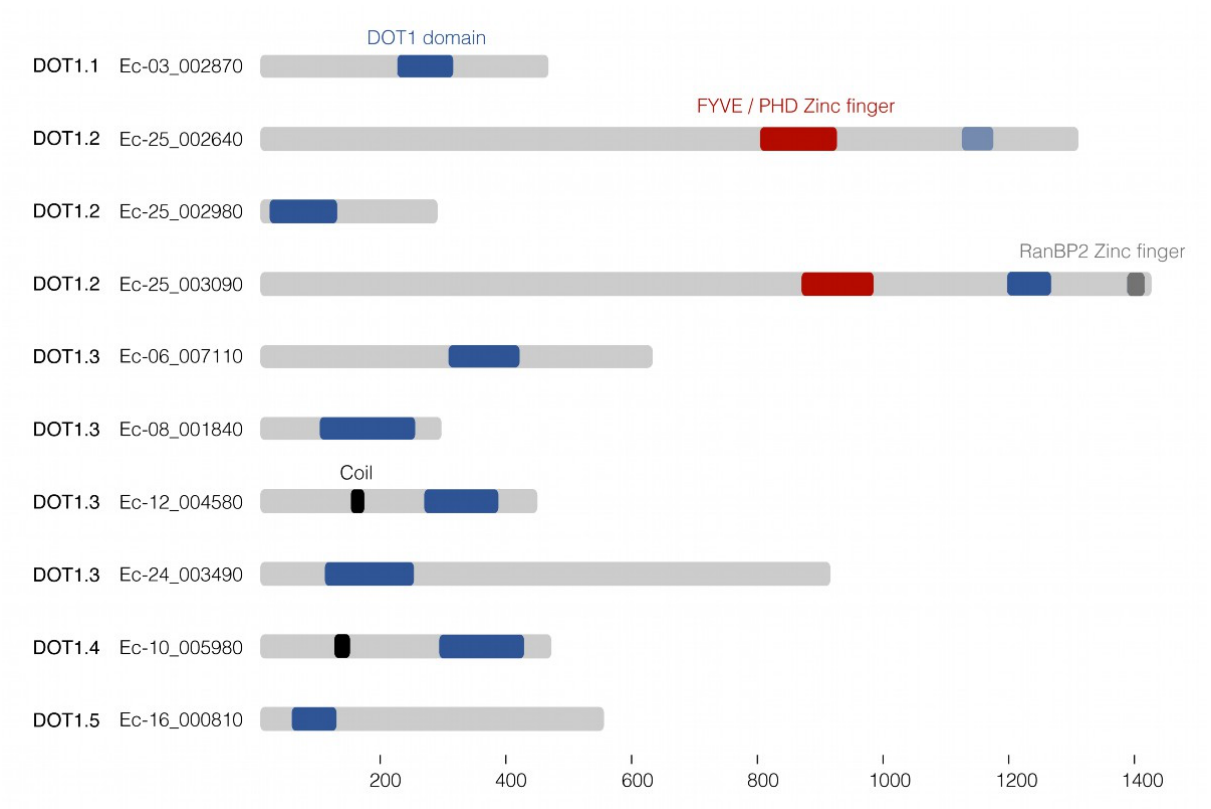
SMC, MB conceived and designed the experiments. JV performed experiments and the algae culture was supported by RL. JV, JSL, MB, FH analyzed the data. PL provided unpublished data and performed experiments. Project Administration: SMC and MB; Resources: SMC; BL carried out the MS experimental work. DL supervised MS data analysis. SMC and MB wrote the manuscript with input from JV and JSL. All authors read and approved the manuscript.

## Acknowledgements

This study was funded by the Max Planck Society, the European Research Council grant 864038 (SMC), the Bettencourt Foundation (SMC) and the Moore Foundation (SMC). The MS platform at the Curie Institut is supported by Région Ile-de-France (N°EX061034) and ITMO Cancer of Aviesan and INCa supported by funds administered by INSERM (no. 21CQ016-00). The LSMP thanks Patrick Poullet from the bioinformatics Core facility (CUBIC) of the Institut Curie U1331 for the continuous development of myProMS. JV and JSL are thankful to the IMPRS 'From molecules to organisms'.

744 **Supplemental Figures**

745




746

747 **Fig. S1**: *Ectocarpus* encodes ten DOT1-domain containing proteins that form part of five distinct gene families.

748

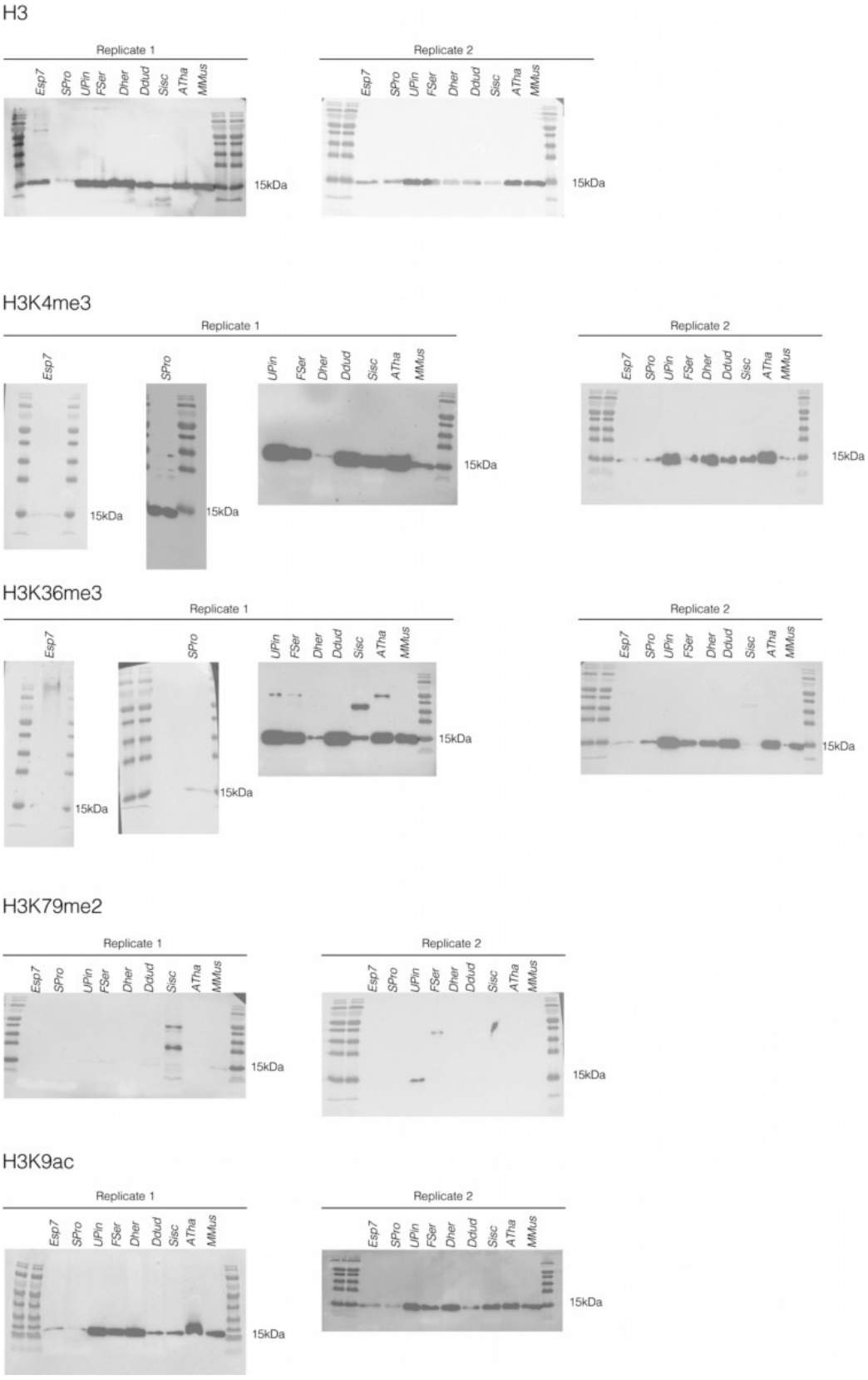
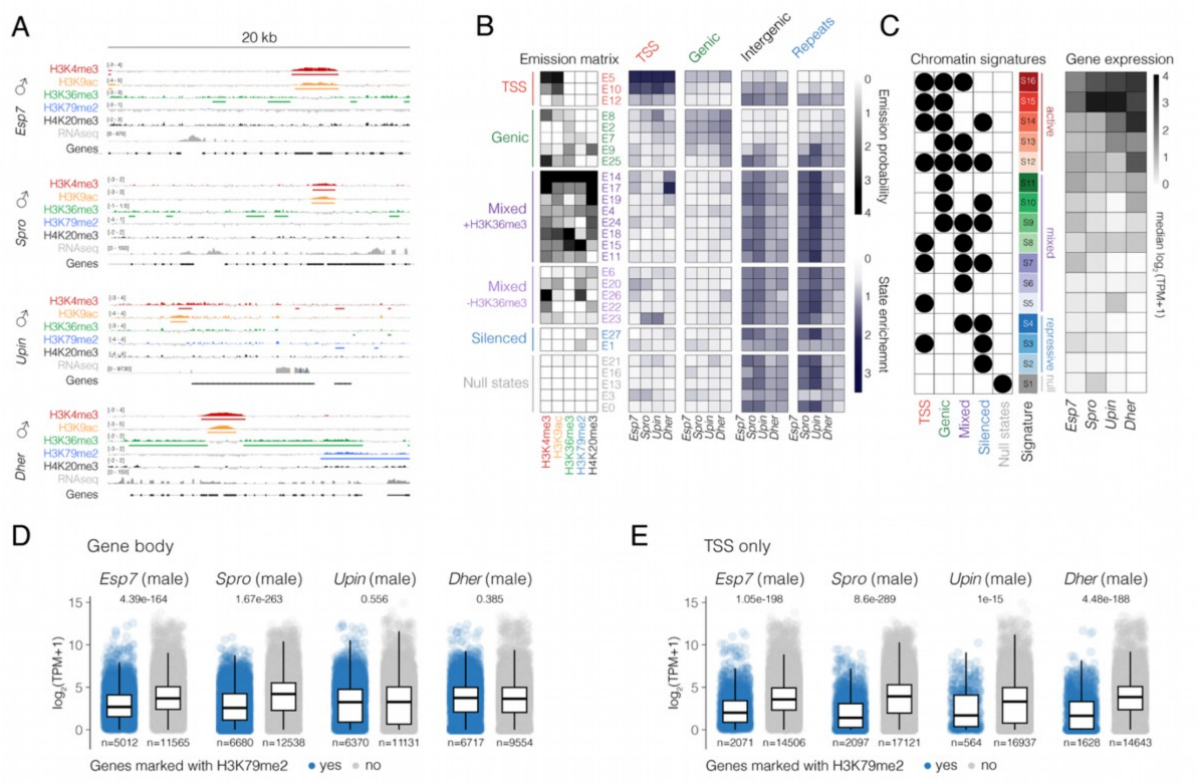


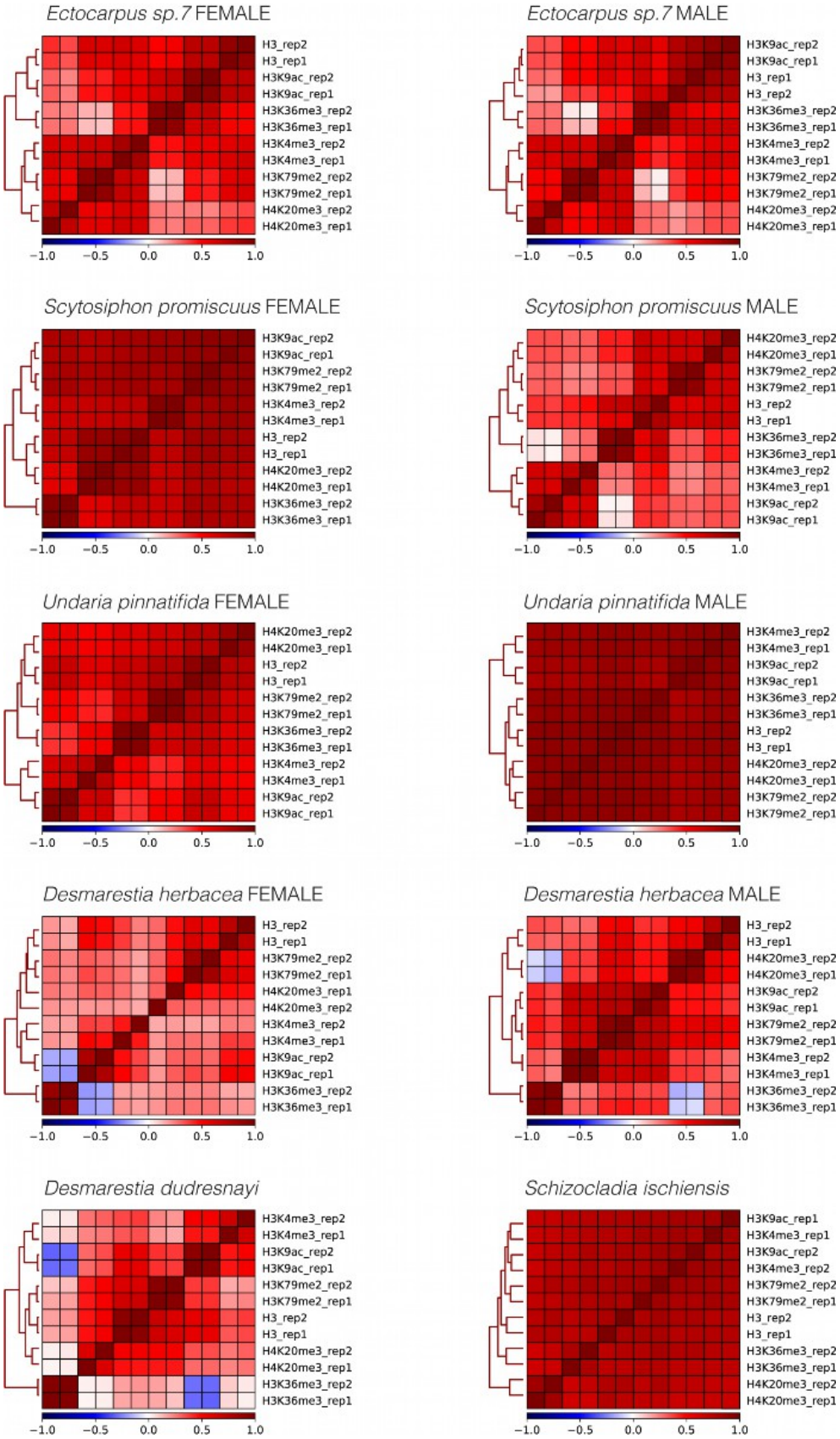

**Fig. S2**: Uncropped Western blots of selected hPTMs across species and model organisms such as *Arabidopsis thaliana* (*ATha*) and *Mus musculus* (*MMus*).

753




754 **Fig. S3**: (**A**) Representative genome browser tracks of ChIP-seq and RNA-seq datasets in male datasets of the
755 dioicous species. ChIP-seq coverage is represented as the $\log_2$ ratio of IP DNA relative to histone H3, with the
756 range indicated on each track. ChIP-seq peaks for each hPTM are indicated under their respective track. (**B**)
757 Enrichment of each chromatin state in genomic features in the male samples of the four dioicous species. (**C**)
758 Median RNA-seq expression level ($\log_2$ TPM+1) of genes associated with each signature in the male samples of
759 the four dioicous species. (**D-E**) Gene expression in males for genes with (blue) or without (grey) an H3K79me2
760 peak overlapping their gene body (**D**) or TSS (**E**).

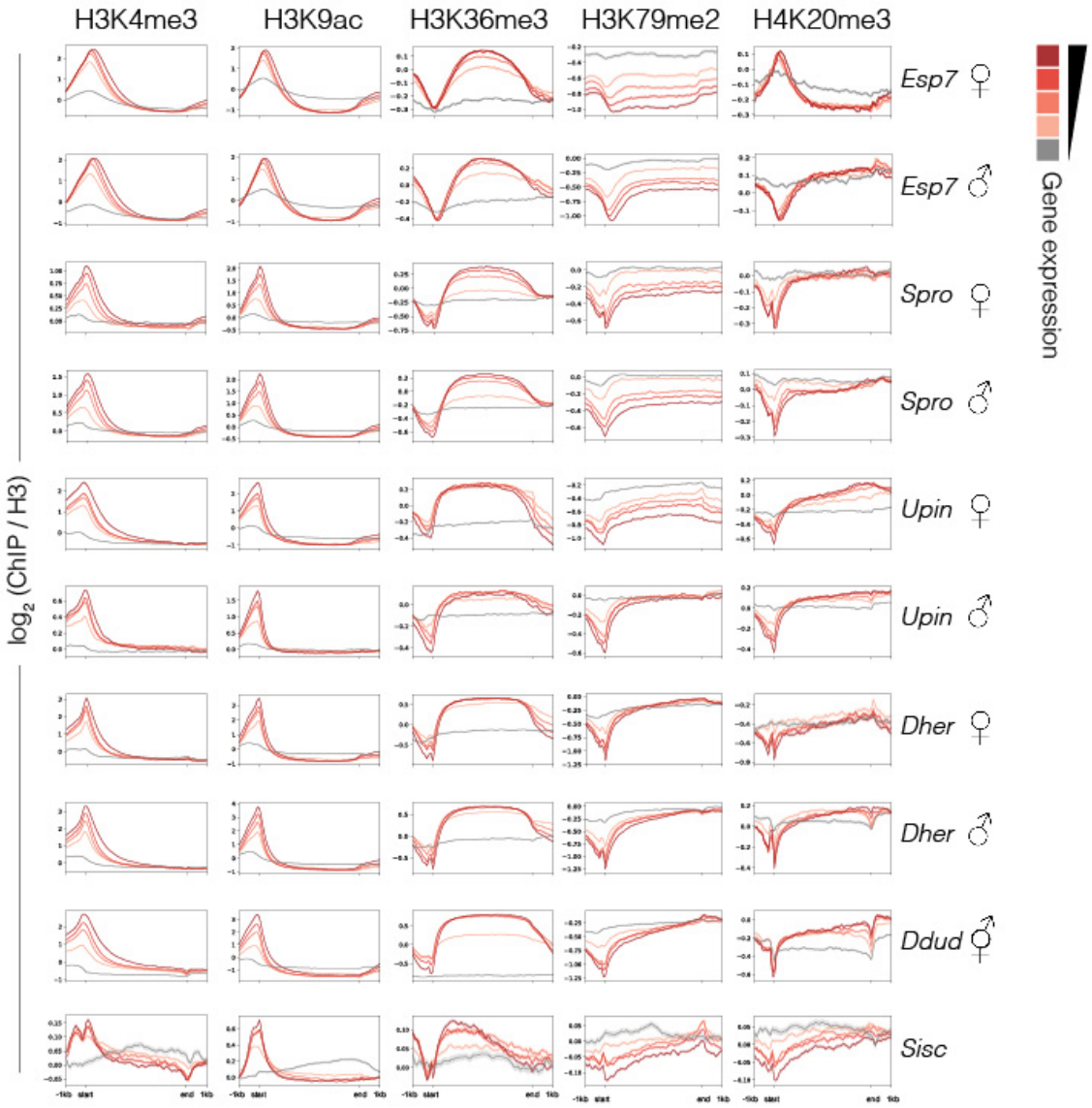
761




762

763 **Fig. S4**: Spearman correlation matrix of the ChIP-seq datasets for each species.

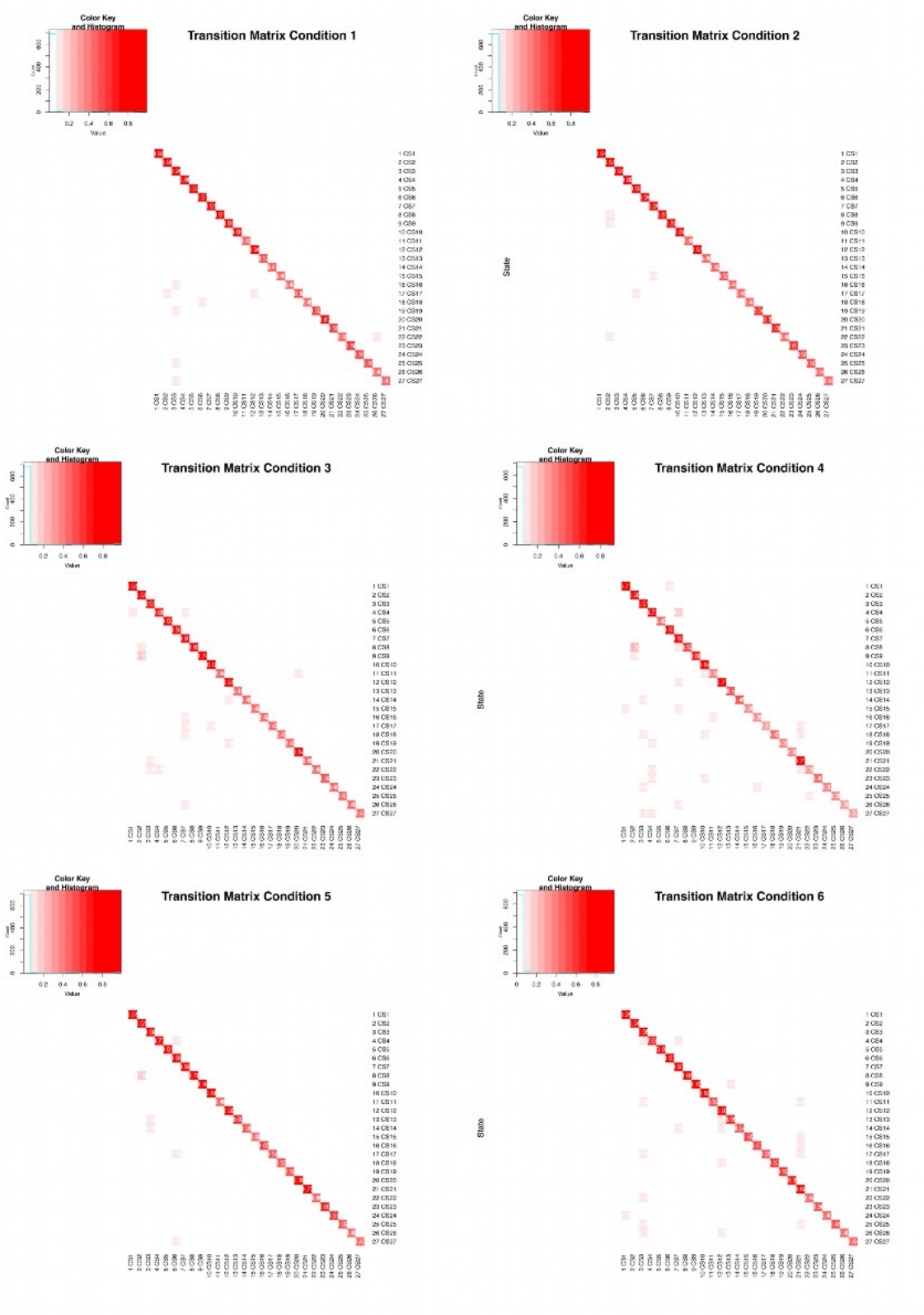
764




765 **Fig. S5**: Metaplots of H3K4me3, H3K9ac, H3K36me3, H3K79me2 and H4K20me3 coverage over genes grouped
766 by gene expression levels in each sample. Gene bodies are plotted as proportional lengths, upstream and
767 downstream intergenic regions in kilobases.

768

769




770 **Fig. S6**: Transition matrices of hiHMM model.

771

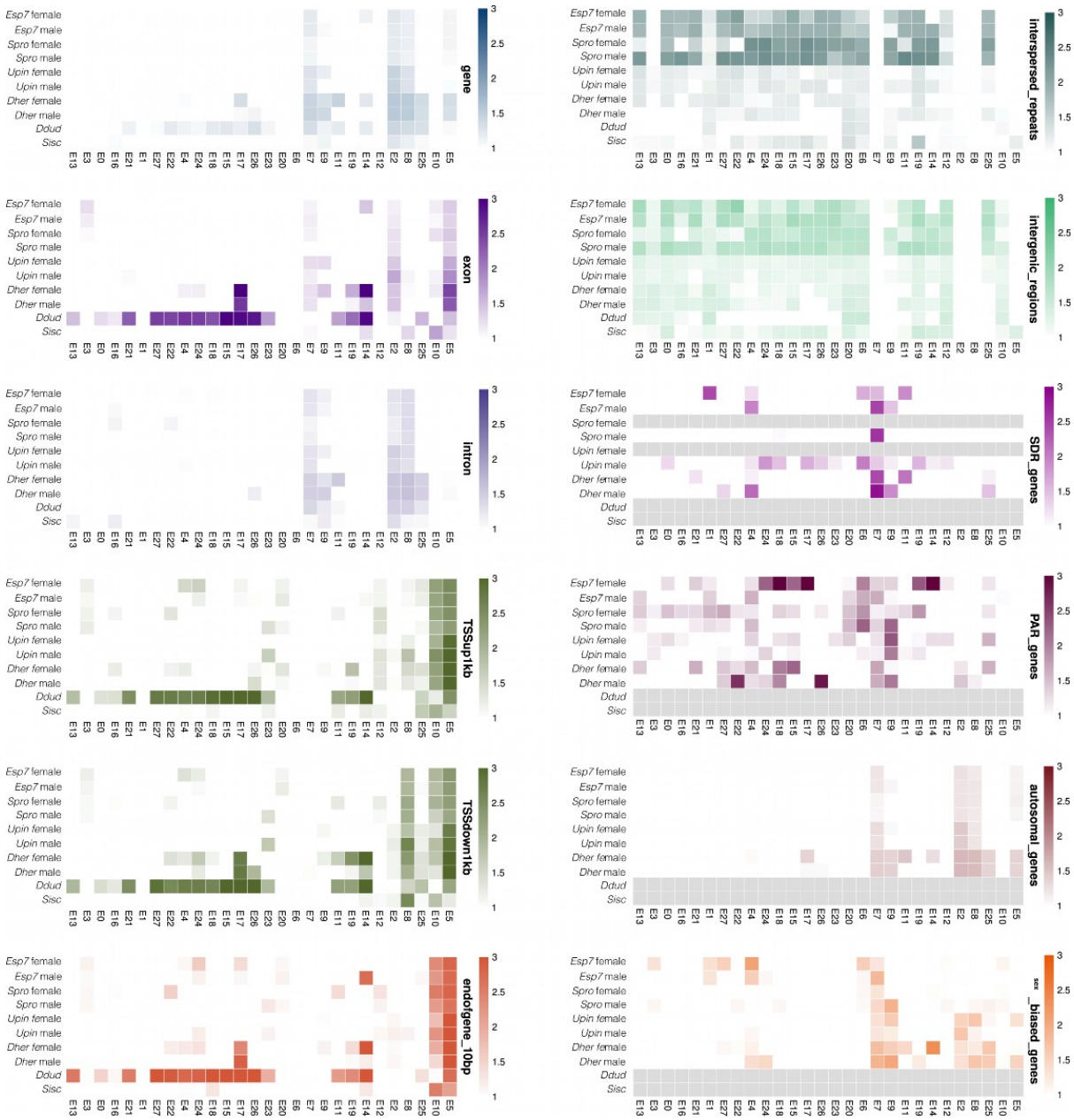
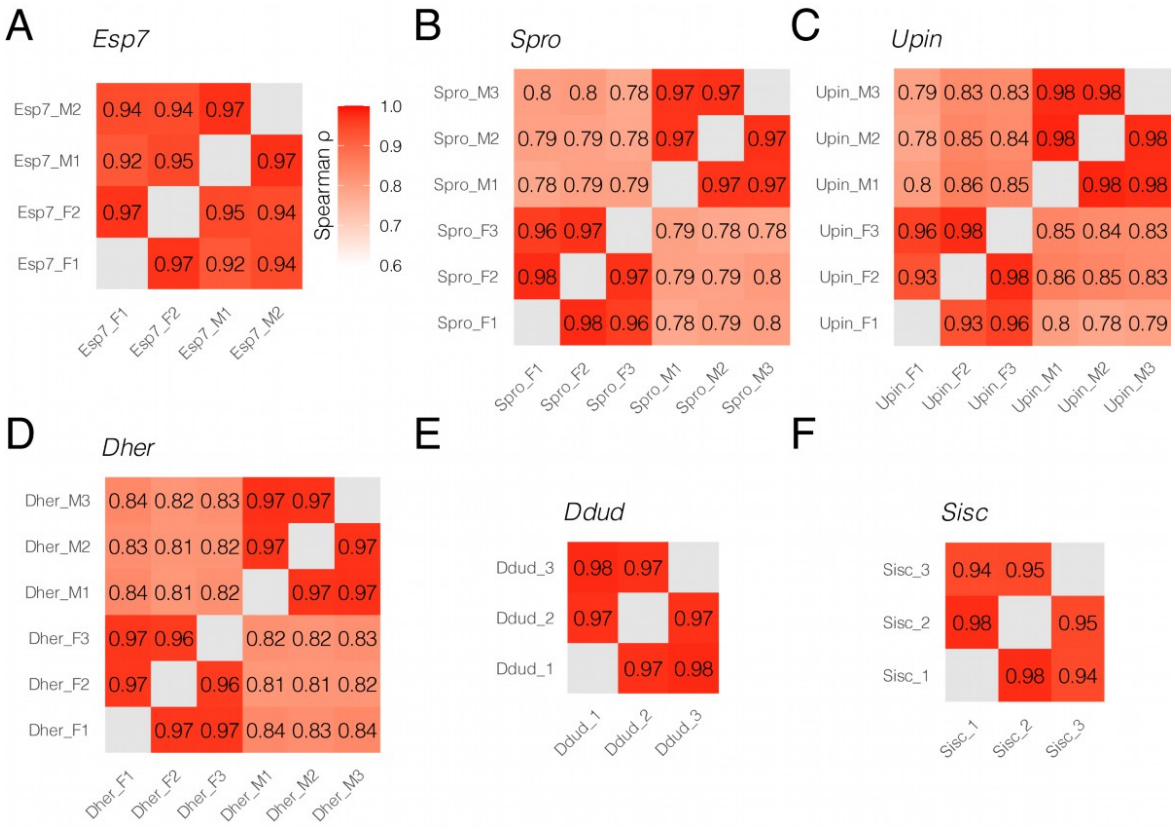



**Fig. S7**: Enrichment of hiHMM emission states over genomic features.

776




777

778 **Fig. S8**: Cross-correlation matrices for the RNAseq datasets in this study. (**A**) *Ectocarpus* (from[8] PRJNA671807),
779 (**B**) *Scytosiphon promiscuus*, (**C**) *Undaria pinnatifida*, (**D**) *Desmarestia herbacea*, (**E**) *Desmarestia dudresnayi*,
780 and (**F**) *Schizocladia ischiensis*. Male and female samples are indicated by suffix "_M" or "_F", respectively.
781 Correlation is measured using Spearman's ρ.

782

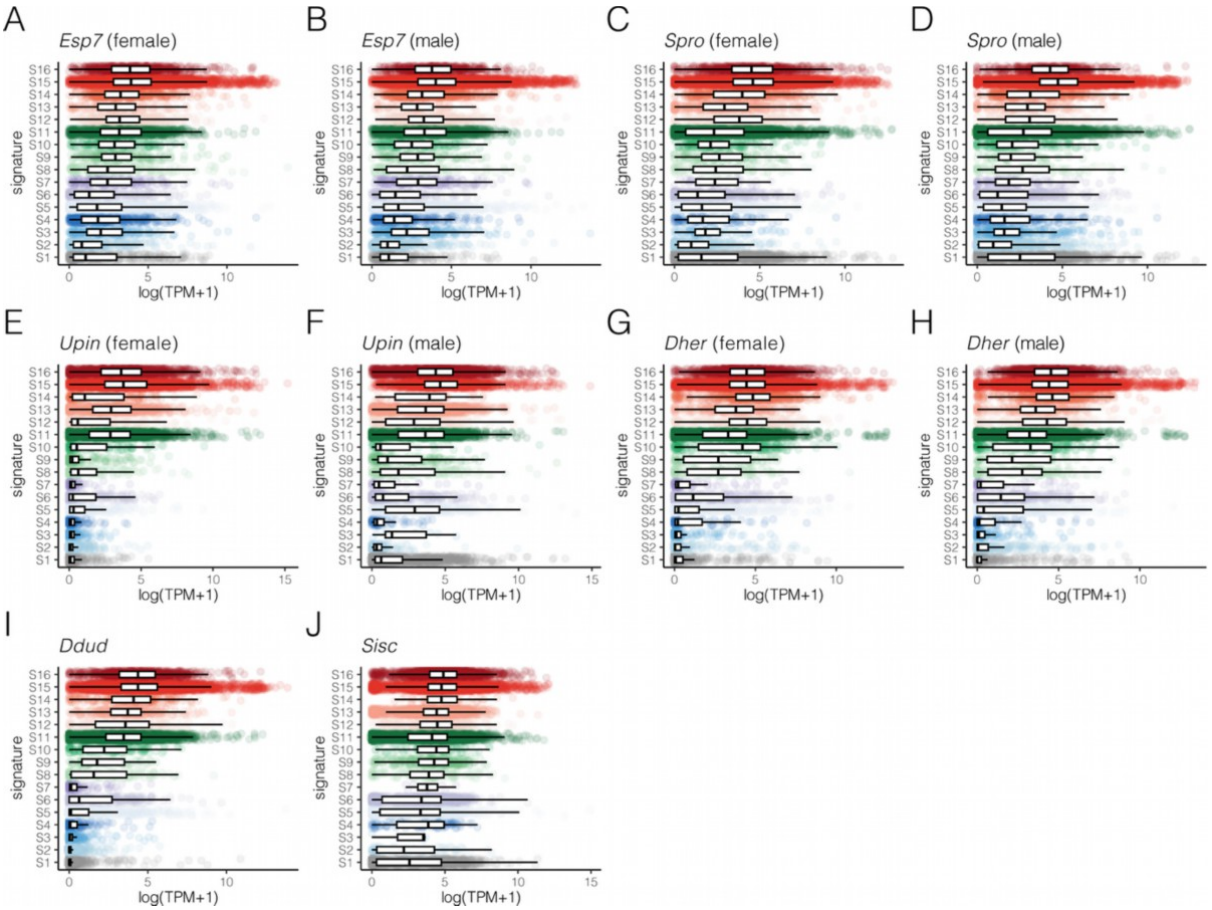



**Fig. S9**: Gene expression levels of genes group by their assigned chromatin signature across all ChIP-seq datasets generated in this study. (**A**) *Ectocarpus* sp 7 female, (**B**) *Ectocarpus* sp 7 male, (**C**) *Scytosiphon promiscuus* female, (**D**) *S. promiscuus* male, (**E**) *Undaria pinnatifida* female, (**F**) *U. pinnatifida* male, (**G**) *Desmarestia herbacea* female, (**H**) *D. herbacea* (male), (**I**) *Desmarestia dudresnayi*, and (**J**) *Schizocladia ischiensis*.

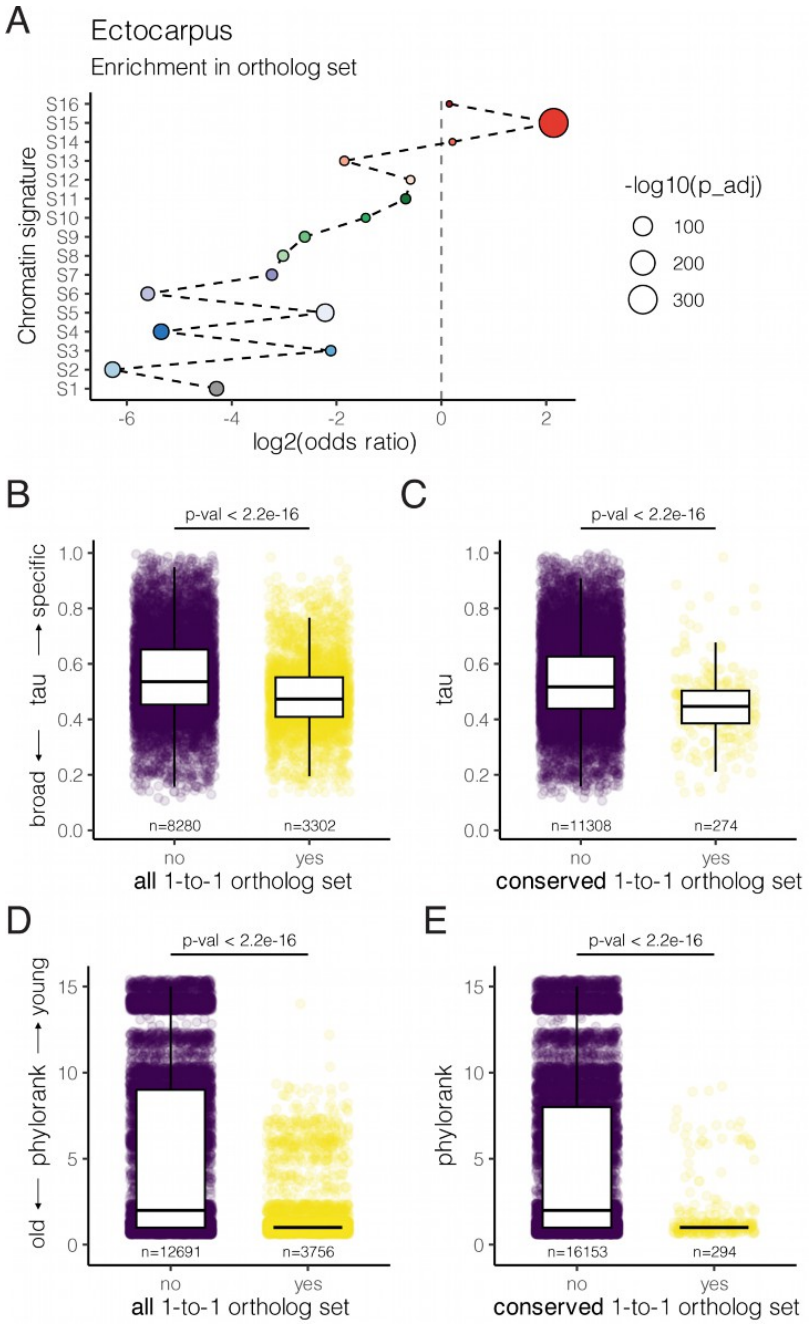



**Fig. S10**: Expression and evolutionary profile of genes in the one-to-one ortholog set compared to all other genes in *Ectocarpus*. (**A**) Chromatin signatures enriched in the one-to-one ortholog set. The Benjamini-Hochberg Procedure was used to adjust the p-values computed from Fisher's exact test. (**B**) Distribution of expression specificity score in all genes on the one-to-one ortholog set and (**C**) in a subset of one-to-one orthologs with conserved chromatin signature across all species. (**D**) Distribution of gene age (phylorank) in all genes on the one-to-one ortholog set and (**E**) in a subset of one-to-one orthologs with conserved chromatin signature across all species. Related to Fig. 3A. The Kruskal–Wallis test is used to compute the p-values in (**B**-**E**).

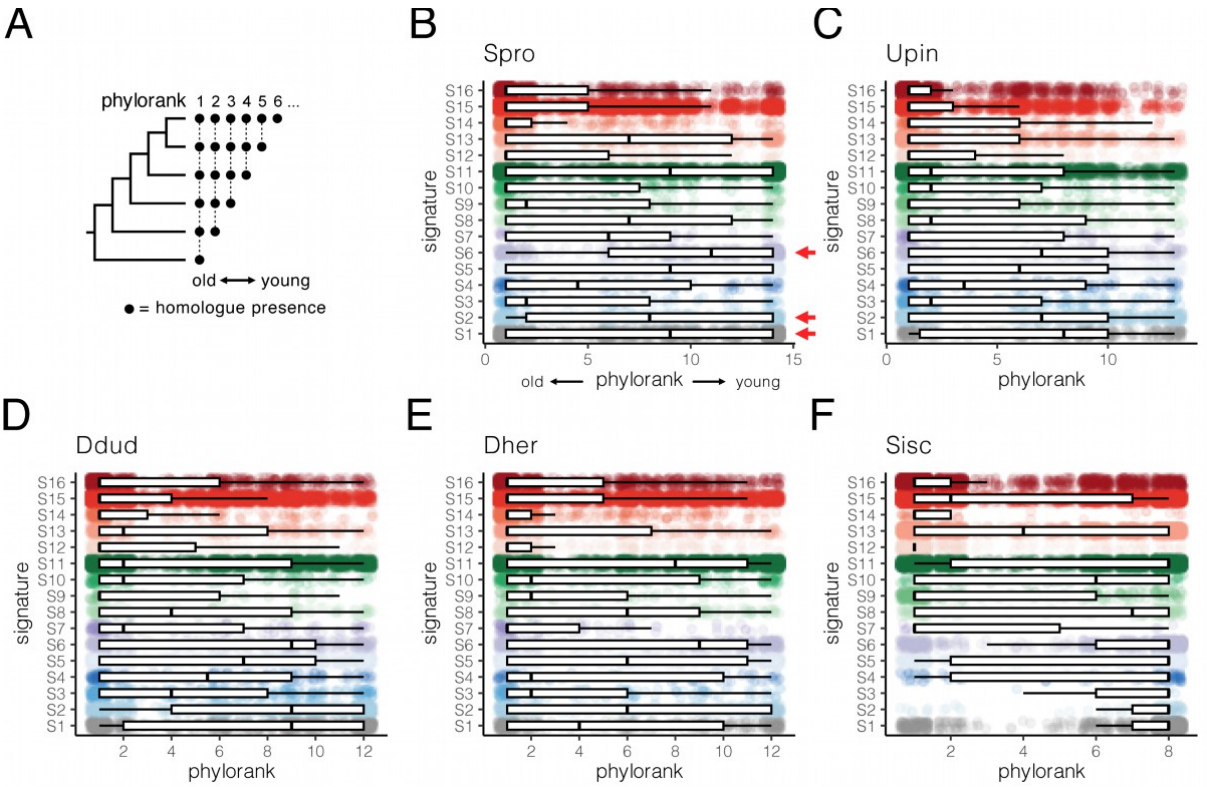
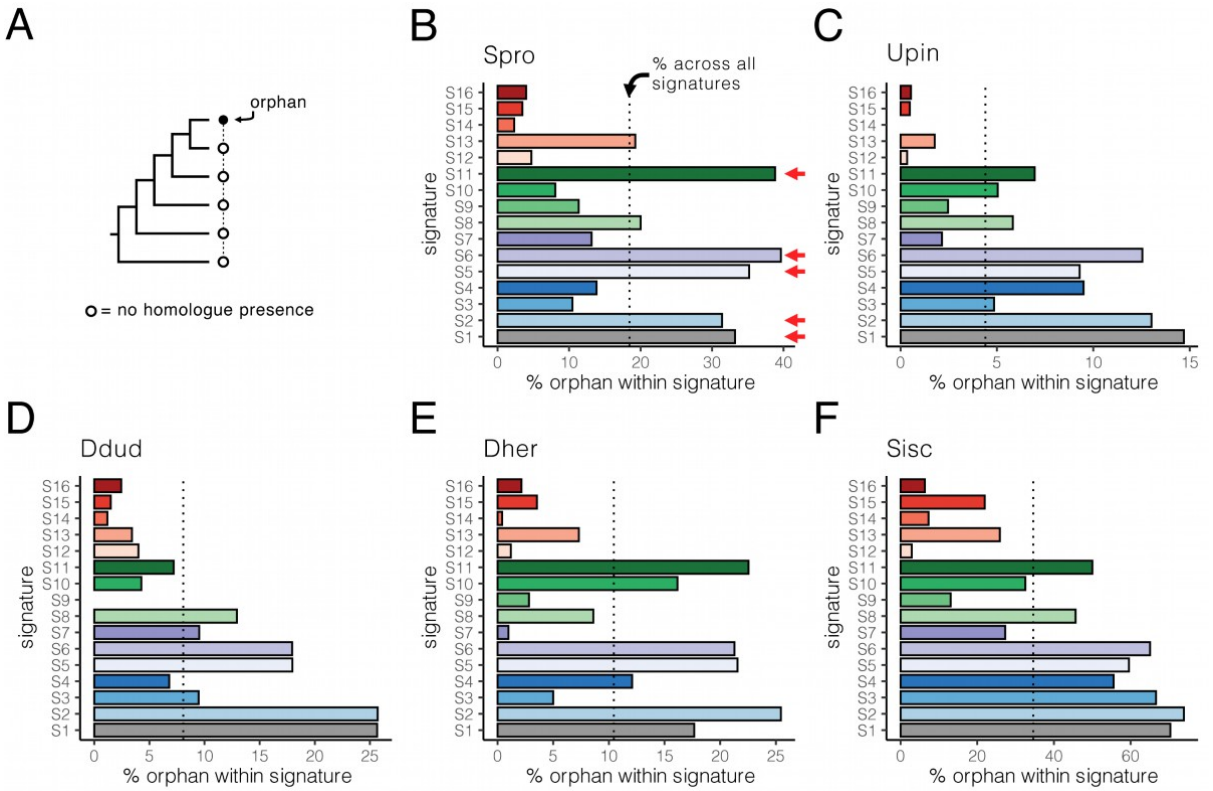



**Fig. S11**: Distribution of gene age across each chromatin signature. (**A**) Summary of the approach to infer gene age (phylorank) via pairwise sequence alignment. The resulting gene age distribution in each chromatin signature for (**B**) *Scytosiphon promiscuus*, (**C**) *Undaria pinnatifida*, (**D**) *Desmarestia dudresnayi*, (**E**) *Desmarestia herbacea*, and (**F**) *Schizocladia ischiensis*. The red arrows mark the chromatin signatures with consistently high distribution of young genes across species. Related to Fig. 3C. Female chromatin signatures were used for dioicous species.

809




810 **Fig. S12**: Percentage of orphan genes assigned to each chromatin signature. (**A**) Summary of the approach to infer
811 orphan genes via pairwise sequence alignment. The resulting percentage of orphan genes detected in each
812 chromatin signature for (**B**) *Scytosiphon promiscuus*, (**C**) *Undaria pinnatifida*, (**D**) *Desmarestia dudresnayi*, (**E**)
813 *Desmarestia herbacea*, and (**F**) *Schizocladia ischiensis*. The red arrows mark the chromatin signatures with
814 consistently high orphan gene presence across species. Related to Fig. 3D. Female chromatin signatures were used
815 for dioicous species.

816

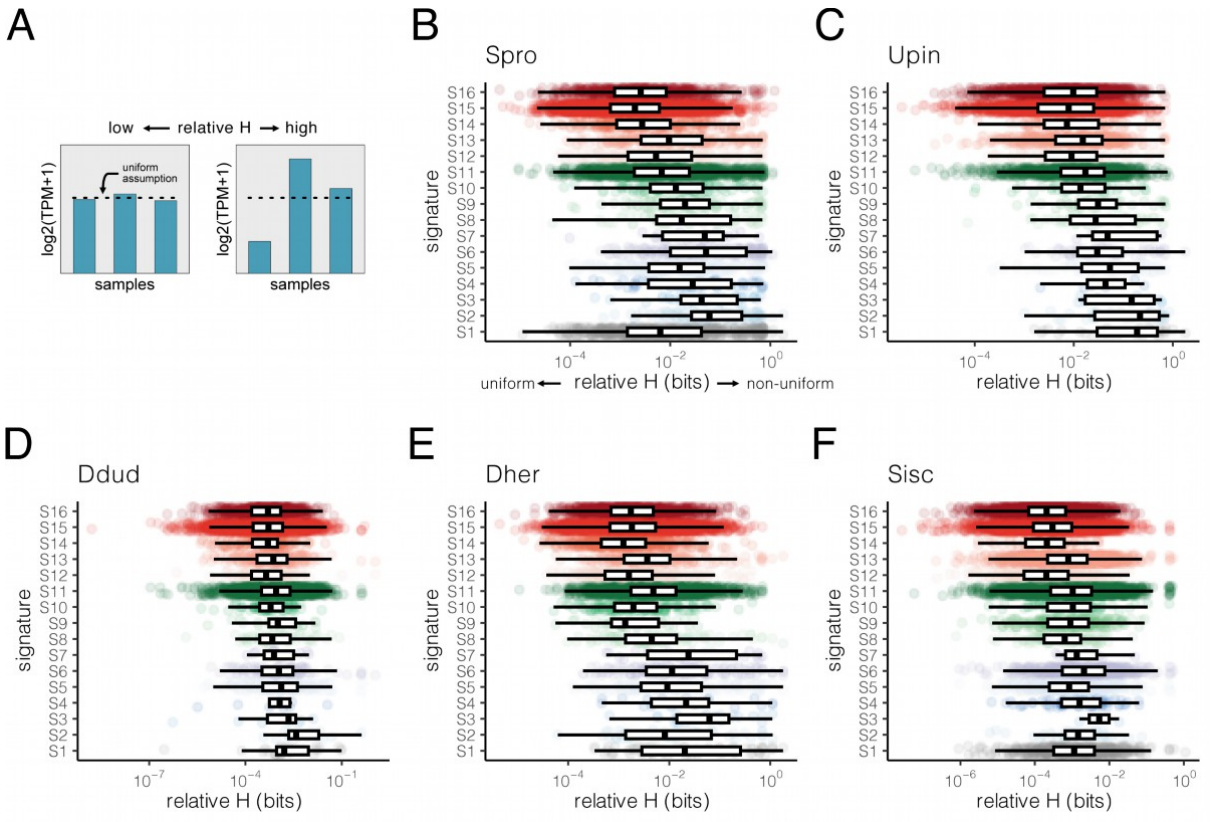
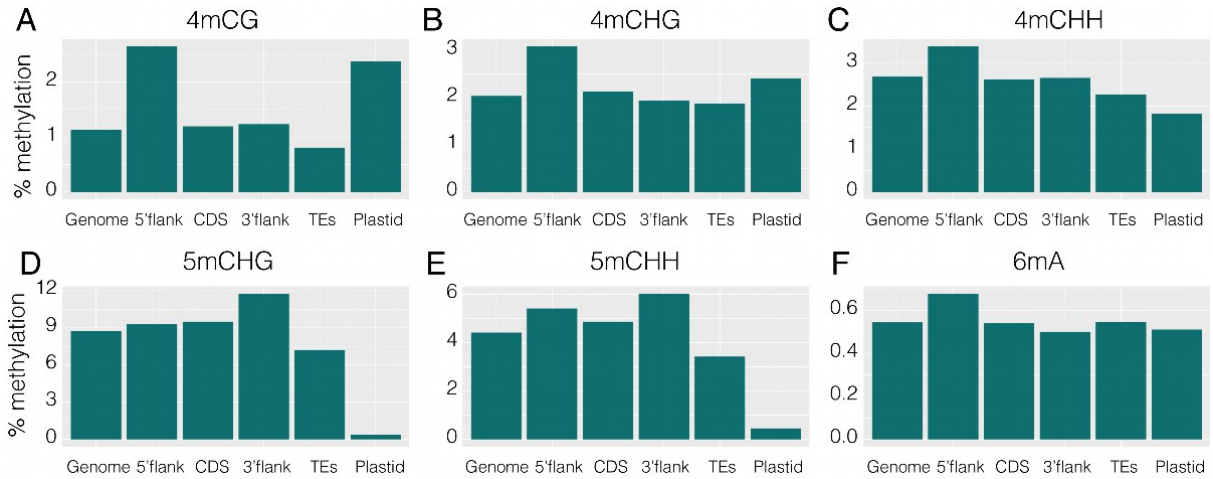



**Fig. S13**: Distribution of expression uniformity across RNA-seq libraries for genes assigned to each chromatin signature. (**A**) Summary of the relative entropy (relative H) approach to compare the observed $\log_2$(TPM+1) expression values across replicates and a gene-specific null assumption of uniform expression values. The resulting relative H scores across chromatin signature for (**B**) *Scytosiphon promiscuus*, (**C**) *Undaria pinnatifida*, (**D**) *Desmarestia dudresnayi*, (**E**) *Desmarestia herbacea*, and (**F**) *Schizocladia ischiensis*. Related to Fig. 3E.




825

826 **Fig. S14**: Average genome-wide levels of (**A-**C) 4mC, (**D-**E) 5mC and (**F**) 6mA in all relevant contexts at different
827 genomic features of the nuclear genome and plastid genome in *S. ischiensis.*

828

829

## Supplemental Table Legends

Table S1 – Species used, strain and genome assembly reference.

Table S2 – Summary of mass spectrometry results.

Table S3 - Quality of the protein samples for mass spectrometry analysis.

Table S4 - ChIP-seq quality metrics.

Table S5 - Data table for *Ectocarpus* genes.

Table S6 - Data table for *Scytosiphon promiscuus* genes.

Table S7 - Data table for *Undaria pinnatifida* genes.

Table S8 - Data table for *Desmarestia herbacea* genes.

Table S9 - Data table for *Desmarestia dudresnayi* genes.

Table S10 - Data table for *Schizocladia ischiensis* genes.

Table S11 - Orthogroups across species.

Table S12 - Enrichment and statistics of signature switches between sexes in brown algae.

Table S13 - List of chromosomes used to train hiHMM model and specific command lines.

## References

1. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).

2. Millán-Zambrano, G., Burton, A., Bannister, A. J. & Schneider, R. Histone post-translational modifications — cause and consequence of genome function. *Nat Rev Genet* **23**, 563–580 (2022).

3. Szalay, M.-F., Majchrzycka, B., Jerković, I., Cavalli, G. & Ibrahim, D. M. Evolution and function of chromatin domains across the tree of life. *Nat Struct Mol Biol* **31**, 1824–1837 (2024).

4. Reinberg, D. & Vales, L. D. Chromatin domains rich in inheritance. *Science* **361**, 33–34 (2018).

5. Deniz, Ö., Frost, J. M. & Branco, M. R. Regulation of transposable elements by DNA modifications. *Nat Rev Genet* **20**, 417–431 (2019).

858 6. Feng, S., Jacobsen, S. E. & Reik, W. Epigenetic Reprogramming in Plant and Animal
859 Development. *Science* **330**, 622–627 (2010).

860 7. Navarrete, C., Montgomery, S. A., Mendieta, J., Lara-Astiaso, D. & Sebé-Pedrós, A.
861 Diversity and evolution of chromatin regulatory states across eukaryotes. Preprint at
862 https://doi.org/10.1101/2025.03.17.643675 (2025).

863 8. Gueno, J. *et al.* Chromatin landscape associated with sexual differentiation in a UV sex
864 determination system. *Nucleic Acids Res* **50**, 3307–3322 (2022).

865 9. Bourdareau, S. *et al.* Histone modifications during the life cycle of the brown alga
866 Ectocarpus. *Genome Biology* **22**, 12 (2021).

867 10. Vigneau, J. *et al.* Interactions between U and V sex chromosomes during the life cycle
868 of *Ectocarpus*. *Development* dev.202677 (2024) doi:10.1242/dev.202677.

869 11. Vigneau, J. & Borg, M. The epigenetic origin of life history transitions in plants and
870 algae. *Plant Reprod* **34**, 267–285 (2021).

871 12. Denoeud, F. *et al.* Evolutionary genomics of the emergence of brown algae as key
872 components of coastal ecosystems. *Cell* S0092867424012728 (2024)
873 doi:10.1016/j.cell.2024.10.049.

874 13. Choi, S.-W. *et al.* Ordovician origin and subsequent diversification of the brown algae.
875 *Current Biology* **34**, 740-754.e4 (2024).

876 14. Heesch, S. *et al.* Evolution of life cycles and reproductive traits: insights from the brown
877 algae. *Journal of Evolutionary Biology* **n/a**, (2021).

878 15. Barrera-Redondo, J. *et al.* Origin and evolutionary trajectories of brown algal sex
879 chromosomes. *Nat Ecol Evol* (2025) doi:10.1038/s41559-025-02838-w.

880 16. Gil, J. & O'Loghlen, A. PRC1 complex diversity: where is it taking us? *Trends in Cell*

*Biology* **24**, 632–641 (2014).

17. Nguyen, A. T. & Zhang, Y. The diverse functions of Dot1 and H3K79 methylation. *Genes Dev.* **25**, 1345–1358 (2011).

18. Cossard, G. G., Toups, M. A. & Pannell, J. R. Sexual dimorphism and rapid turnover in gene expression in pre-reproductive seedlings of a dioecious herb. *Annals of Botany* mcy183–mcy183 (2018) doi:10.1093/aob/mcy183.

19. Liesner, D. *et al.* Developmental pathways underlying sexual differentiation in the U/V sex chromosome system of giant kelp. *Developmental Cell* S1534580724007615 (2025) doi:10.1016/j.devcel.2024.12.022.

20. Kawai, H., Maeba, S., Sasaki, H., Okuda, K. & Henry, E. C. Schizocladia ischiensis: A New Filamentous Marine Chromophyte Belonging to a New Class, Schizocladiophyceae. *Protist* **154**, 211–228 (2003).

21. Sohn, K.-A. *et al.* hiHMM: Bayesian non-parametric joint inference of chromatin state maps. *Bioinformatics* **31**, 2066–2074 (2015).

22. Barrera-Redondo, J. *et al. Origin and Evolutionary Trajectories of Brown Algal Sex Chromosomes*. http://biorxiv.org/lookup/doi/10.1101/2024.01.15.575685 (2024) doi:10.1101/2024.01.15.575685.

23. Barrera-Redondo, J., Lotharukpong, J. S., Drost, H.-G. & Coelho, S. M. Uncovering gene-family founder events during major evolutionary transitions in animals, plants and fungi using GenEra. *Genome Biol* **24**, 54 (2023).

24. Bachtrog, D. *et al.* Sex determination: why so many ways of doing it? *PLoS Biol* **12**, e1001899 (2014).

25. Coelho, S. M., Gueno, J., Lipinska, A. P., Cock, J. M. & Umen, J. G. UV chromosomes

904 and haploid sexual systems. *Trends Plant Sci* **23**, 794–807 (2018).

905 26. Illingworth, R. S. & Bird, A. P. CpG islands – 'A rough guide'. *FEBS Letters* **583**,
906 1713–1720 (2009).

907 27. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.*
908 **25**, 1010–1022 (2011).

909 28. Schmitz, R. J., Lewis, Z. A. & Goll, M. G. DNA Methylation: Shared and Divergent
910 Features across Eukaryotes. *Trends in Genetics* **35**, 818–827 (2019).

911 29. Wiles, E. T. & Selker, E. U. H3K27 methylation: a promiscuous repressive chromatin
912 mark. *Current Opinion in Genetics & Development* **43**, 31–37 (2017).

913 30. Zarif, M., Rousselot, E., Jesus, B., Tirichine, L. & Duc, C. H3K27me3 and EZH Are
914 Involved in the Control of the Heat-Stress-Elicited Morphological Changes in Diatoms.
915 *IJMS* **25**, 8373 (2024).

916 31. Malla, A. B. *et al.* DOT1L bridges transcription and heterochromatin formation at
917 mammalian pericentromeres. *EMBO Reports* **24**, e56492 (2023).

918 32. Verzijlbergen, K. F., Faber, A. W., Stulemeijer, I. J. & Van Leeuwen, F. Multiple
919 histone modifications in euchromatin promote heterochromatin formation by redundant
920 mechanisms in Saccharomyces cerevisiae. *BMC Molecular Biol* **10**, 76 (2009).

921 33. Zhang, J.-Y. & Zhou, Q. On the Regulatory Evolution of New Genes Throughout Their
922 Life History. *Molecular Biology and Evolution* **36**, 15–27 (2019).

923 34. Werner, M. S. *et al.* Young genes have distinct gene structure, epigenetic profiles, and
924 transcriptional regulation. *Genome Research* **28**, 1675–1687 (2018).

925 35. Petroll, R. *et al.* The Expansion and Diversification of Epigenetic Regulatory Networks
926 Underpins Major Transitions in the Evolution of Land Plants. *Molecular Biology and*

927 *Evolution* **42**, msaf064 (2025).

928 36. Cui, X. *et al.* Young Genes out of the Male: An Insight from Evolutionary Age Analysis
929 of the Pollen Transcriptome. *Mol Plant* **8**, 935–945 (2015).

930 37. Begun, D. J., Lindfors, H. A., Kern, A. D. & Jones, C. D. Evidence for *de Novo*
931 Evolution of Testis-Expressed Genes in the *Drosophila yakuba* / *Drosophila erecta* Clade.
932 *Genetics* **176**, 1131–1137 (2007).

933 38. Lipinska, A. P. *et al.* Multiple gene movements into and out of haploid sex
934 chromosomes. *Genome Biology* **18**, 104 (2017).

935 39. Ahmed, S. *et al.* A haploid system of sex determination in the brown alga *Ectocarpus*
936 sp. *Curr. Biol.* **24**, 1945–1957 (2014).

937 40. Avia, K. *et al.* Genetic diversity in the UV sex chromosomes of the brown alga
938 *Ectocarpus*. *Genes (Basel)* **9**, (2018).

939 41. Oldroyd, B. P. & Yagound, B. The role of epigenetics, particularly DNA methylation,
940 in the evolution of caste in insect societies. *Phil. Trans. R. Soc. B* **376**, 20200115 (2021).

941 42. Muyle, A., Bachtrog, D., Marais, G. A. B. & Turner, J. M. A. Epigenetics drive the
942 evolution of sex chromosomes in animals and plants. *Phil. Trans. R. Soc. B* **376**, 20200124
943 (2021).

944 43. Zhou, Q. *et al.* The epigenome of evolving Drosophila neo-sex chromosomes: dosage
945 compensation and heterochromatin formation. *PLoS Biol* **11**, e1001711 (2013).

946 44. Namekawa, S. H., VandeBerg, J. L., McCarrey, J. R. & Lee, J. T. Sex chromosome
947 silencing in the marsupial male germ line. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9730–9735
948 (2007).

949 45. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local

950 alignment search tool. *Journal of Molecular Biology* **215**, 403–410 (1990).

951 46. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for
952 comparative genomics. *Genome Biol* **20**, 238 (2019).

953 47. Luthringer, R. *et al.* Repeated co-option of HMG-box genes for sex determination in
954 brown algae and animals. *Science* **383**, eadk5466 (2024).

955 48. Brosch, M., Yu, L., Hubbard, T. & Choudhary, J. Accurate and Sensitive Peptide
956 Identification with Mascot Percolator. *J. Proteome Res.* **8**, 3176–3181 (2009).

957 49. Poullet, P., Carpentier, S. & Barillot, E. ***my*** ProMS, a web server for management and
958 validation of mass spectrometry-based proteomic data. *Proteomics* **7**, 2553–2556 (2007).

959 50. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics
960 pipelines. *Nat Biotechnol* **38**, 276–278 (2020).

961 51. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible
962 platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191 (2014).

963 52. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and
964 dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

965 53. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with
966 ChromHMM. *Nat Protoc* **12**, 2478–2492 (2017).

967 54. Roces, V., Guerrero, S., Álvarez, A., Pascual, J. & Meijón, M. PlantFUNCO:
968 Integrative Functional Genomics Database Reveals Clues into Duplicates Divergence
969 Evolution. *Molecular Biology and Evolution* **41**, msae042 (2024).

970 55. Lotharukpong, J. S. *et al.* A transcriptomic hourglass in brown algae. *Nature* **635**, 129–
971 135 (2024).

972 56. Yanai, I. *et al.* Genome-wide midrange transcription profiles reveal expression level

973 relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005).

974 57. Wickham, H. *et al.* Welcome to the Tidyverse. *JOSS* **4**, 1686 (2019).

975 58. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing

976 with genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).

977

# Lineage-wide evolution of 3D genome organisation and centromeres in brown algae

Pengfei Liu[1], Rory J. Craig[1,3], Elena Avdievich[1], Fabian B. Haas[1], Chang Liu[2], Susana M Coelho[1*]

[1]Department of Algal Development and Evolution, Max Planck Institute for Biology, Max-Planck-Ring 5, 72076 Tübingen, Germany. [2]Institute of Biology, University of Hohenheim, Stuttgart, Germany. [3]Current address: School of BioSciences, University of Melbourne, Parkville, VIC 3010, Australia

[*]Correspondence: susana.coelho@tuebingen.mpg.de

## Abstract

Although 3D genome architecture has been described for an increasing number of plant and algal species, comparative analyses across closely related lineages remain scarce. Consequently, fundamental questions persist about how chromatin organization is maintained or reshaped over deep evolutionary time, and how such changes relate to life-history traits, genome size, and linear genome features. Here, we present a comprehensive analysis of 3D chromatin architecture across six brown algae species and one outgroup, spanning the phylogenetic breadth and biological complexity of this key photosynthetic lineage. We show that compact genomes lack chromatin domains whereas larger, transposable element-rich genomes of morphologically complex taxa tend to exhibit structured organization including TAD-like domains. We investigate chromatin folding patterns and gene expression over evolutionary time and uncover 3D chromatin features associated with transitions in sexual systems. Moreover, we reconstruct the ancestral brown algal karyotype, revealing deeply conserved macrosynteny and providing a new framework for interpreting chromosome-scale genome dynamics. Finally, we uncover an ancient and highly conserved association between centromeres and chromodomain-encoding retrotransposons, revealing a remarkable example of convergence in centromere–transposon co-evolution between brown algae and angiosperms, and one of the most stable examples of centromere-linked transposable elements known in eukaryotes. Together, our findings elucidate the evolutionary history of 3D chromatin and linear genome architectures across an entire eukaryotic lineage and highlight extreme centromere stability in brown algae, providing a powerful point of comparison with land plants and deepening our understanding of genome evolution in independent multicellular lineages.

**Keywords**: 3D genome, Hi-C, topologically associated domains, ancestral genome, chromosomal reorganizations, brown algae

**One sentence summary**: We present a lineage-wide evolutionary analysis of 3D genome architecture and organization across brown algae, revealing conserved chromosomal features, lineage-specific rearrangements, and long-term co-evolution of centromeres and retrotransposons that together illuminate how nuclear architecture evolves over hundreds of millions of years.

## Introduction

Eukaryotic genomes are organized within the 3D nuclear space, and this spatial architecture is fundamental to chromatin folding and genome regulation[1,2]. Advances in chromatin conformation capture technologies, particularly Hi-C, have transformed our understanding of higher-order genome organization across cell types and species [3–8] revealing conserved and divergent principles of chromosome folding. Current models of nuclear architecture recognize that chromosomes in the cell nucleus are organized as chromosome territories (CTs). At finer scales, genomes are partitioned into A/B compartments, topologically associating domains (TADs), and specific promoter–enhancer loops[1,9] providing a multilayered regulatory framework that links 3D structure to gene activity.

The 3D nuclear architecture has been characterized in model plants[10–13], animals[3,9,14–17], and yeast[4,5,18], yet our understanding of chromatin organization across broader eukaryotic lineages remains limited[19]. Moreover, while an increasing number of studies describe the 3D genome of individual organisms[1,20], comparative studies examining the evolutionary dynamics of 3D genome architecture across related species within a lineage are notably scarce, leaving fundamental questions about the conservation and divergence of higher-order chromatin structure largely unresolved. Specifically, the degree to which 3D genome organization is maintained across species boundaries, and the potential correlations between architectural modifications and key life history traits, genome size variation, and other structural genomic changes, remain poorly understood.

In this context, the brown algae (Phaeophyceae) offer an interesting study system. They are the third most complex multicellular eukaryotic lineage[21] having arisen independently from animal and plant lineages[22], although they are photosynthetic organisms that have been extensively used for comparative studies with plant lineages[23,24]. Brown algae as a group present a range of levels of morphological complexity, sexual system and sexual dimorphism, and lengths of haploid and diploid phase of development[28]. Like land plants, brown algae undergo a haploid–diploid alternation of generations, and many species possess sexual systems reminiscent of bryophytes, with UV sex chromosomes and separate sexes[23,24,29,30], although some have transitioned to hermaphroditism (monoicy). These life-history traits are expected to strongly influence genome evolution and 3D genome architecture, yet direct empirical evidence linking them to chromatin organization remains scarce.

Despite the recent sequencing of 65 genomes of brown algae[26], the 3D architecture of the brown algal nucleus is largely unknwon, except for a recent study using the model brown alga *Ectocarpus* sp.7[31]. *Ectocarpus* is a simple, filamentous brown algae with a compact 200 Mbp[32,33] genome. Hi-C analyses showed that *Ectocarpus* genome can be partitioned into loose or compact structural domains that bear some similarities to those of the mammalian A/B compartments. *Ectocarpus* interphase chromatin exhibits a non-Rabl 3D chromatin conformation, with strong contacts among telomeres and among centromeres, which feature centromere-specific LTR retrotransposons{Citation}. However, *Ectocarpus* chromosomes do not contain large local interactive domains (i.e., TADs), which are a predominant feature of animal genomes[34] and large genomes of plants[35,36].

However, Ectocarpus lacks large local interactive domains (TADs), which are prominent in animals and large plant genomes, leaving open the question of whether its 3D genome structure is representative of the lineage.

Here, we present a comparative analysis of 3D chromatin architecture across six brown algal species plus one outgroup, spanning a range of sexual systems, genome sizes, and morphological complexity. We find that compact genomes lack chromatin domains, whereas larger, transposable element–rich genomes exhibit structured organization, including TAD-like domains, linking genome complexity with 3D architecture. Sex chromosome 3D structure is maintained during transitions to hermaphroditism, reflecting linear conservation of male-specific features and demonstrating long-term stability of chromosome-level architecture. Using new chromosome-scale assemblies, we reconstruct ancestral karyotypes, revealing deeply conserved macrosynteny alongside lineage-specific rearrangements. Finally, we characterize brown algal centromeres, uncovering their long-term co-evolution with centrophilic retrotransposons and remarkable conservation of centromeric regions over ~240 million years. These findings position brown algae as a powerful comparative system for understanding the evolution of 3D genome organization and centromere biology, providing insights highly relevant to plant genomics and evolution.

# Results

## Hi-C-guided chromosome level assemblies reveal conserved macrosynteny

To assess the evolution of chromosome structure across a diverse set of brown algae, we utilized Hi-C-guided, chromosome-level assemblies for six brown algal genomes and one sister outgroup (**Table S1, S2**). These brown algal species represent four major orders, encompassing the phylogenetic history of the lineage[21,37] as well as its diversity in terms of sexual systems, sexual dimorphism, morphological complexity, and types of life cycle[28] (**Fig. 1A-B**). We built on the recently published chromosome-level assembly of *Ectocarpus* sp. 7[31], and improved the assemblies of five recently published algal genomes[25,26] to chromosomal-level using Hi-C (see methods): *Chordaria linearis, Undaria pinnatifida, Desmarestia herbacea, Desmarestia dudresnayi,* and *Dictyota dichotoma*. Furthermore, we produced *de novo* chromosomal-level assemblies using Nanopore and Hi-C technologies for *U. pinnatifida* and *D. dichotoma* female strains (**Fig. 1B, Fig. S1, Table S1, S2**, see methods), enabling comparison between male and female genomes for these dioicous species. The new assemblies show high DNA pairwise alignment identity compared to previous assemblies (**Fig. S2**) and yield genome sizes from 185.8 Mb (*Ectocarpus* sp.7[31]) to 848.5 Mb (*D. dichotoma*), with N50 values from 6.91 Mb to 28.1 Mb and BUSCO scores ranging from 87.1% to 89.0% completeness using the eukaryota_odb10 database[38] (**Fig. S1, Table S2**). The differences in genome size are mainly explained by expansions of transposable elements (TEs) in species with larger genomes (**Fig. 1B, S3**).

We also achieved a near-contiguous chromosome-level assembly for the sister group species *Schizocladia ischiensis* (**Table S2**) which has, as previously suggested, a highly rearranged genome relative to brown algae[25]. For synteny analyses, we focused on orthologs that were present on chromosomes (and excluded unlinked scaffolds, see methods) (**Fig. 1C, Fig. S1, Table S2**). Brown algae have 27-32 chromosomes and largely conserved macrosynteny and gene collinearity even after app. 240 million years of independent evolution (**Fig. 1B-C**), often differing by a few discrete events superimposed on a background rate of small-scale gene transfers between chromosomes. The evolutionary history of brown algal linear genomes is analyzed in greater detail below.

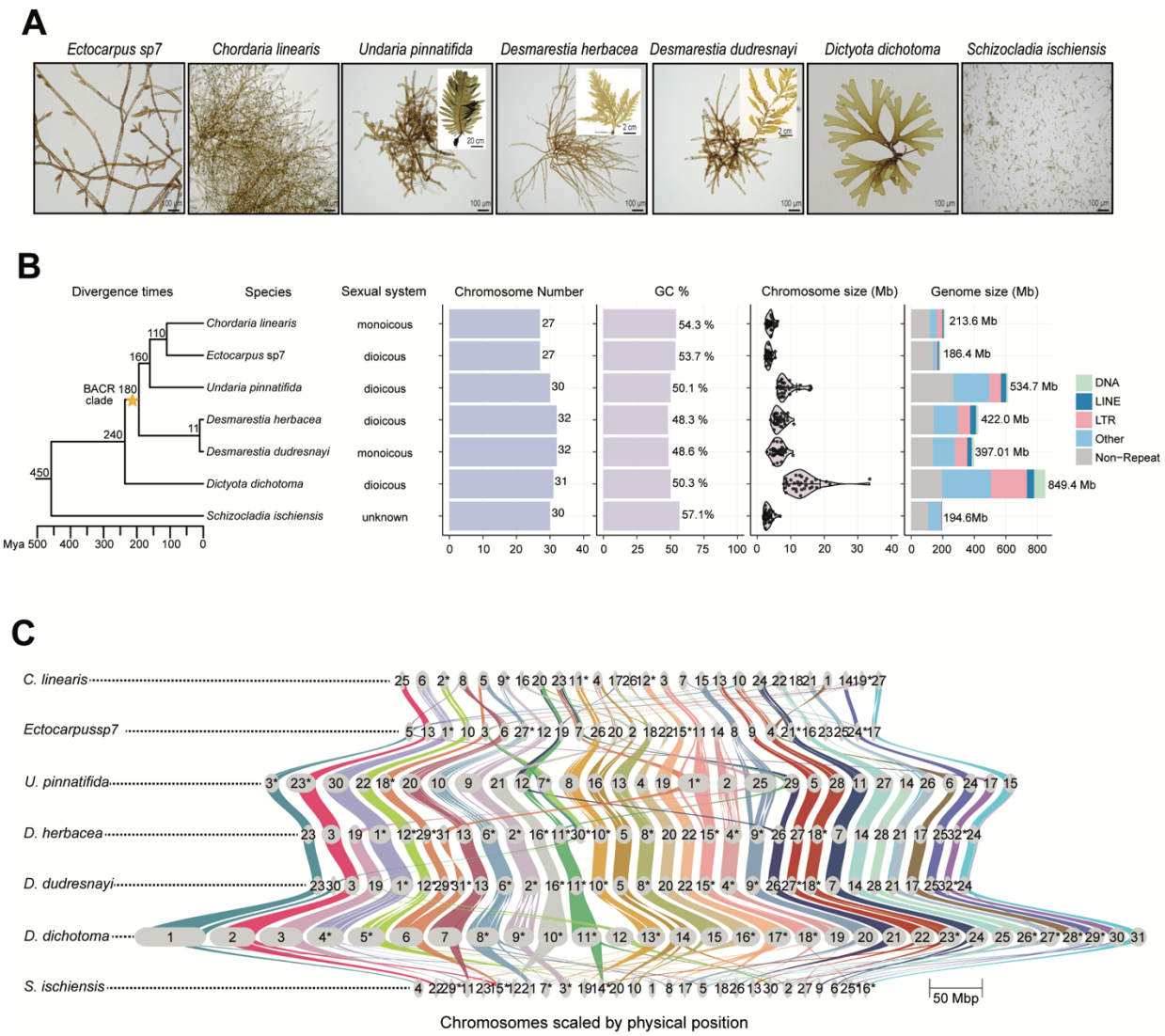



**Fig. 1. High quality genome assemblies for representative brown algal species and outgroup**. **(A)** Species used in this study, covering the phylogenetic and morphological diversity of brown algae. Note that the images shown represent the gametophyte (haploid) generation whereas the inserts represent the morphologically complex sporophyte generation (diploid). **(B)** Phylogenetic position of the studied species and genome statistics. The phylogeny was built in TreeViewerV2.20[39]; the nodes are labelled with the approximate divergence time from[26], the BACR (brown algal crown radiation) clade is marked with golden brown star. Repeats were grouped into major classes (DNA, LTR, LINE)[40]. All remaining categories, including unclassified repeats, simple repeats, satellites, and others, were combined into 'other'. (**C**) Ribbon plot of syntenic orthologous genes conserved among brown algae and outgroup. The colored vertical links connect orthologous genes (identified using GENESPACE[41]) to the numbered chromosomes among the species. The order and color the chromosomes are rephased by *D. dichotoma*, and inverted chromosomes are label with an asterisk.

### 3D genome architecture of brown algae genomes

To uncover how 3D genome organization has evolved in brown algae, we utilized the new Hi-C datasets for the five focal species and one sister outgroup, and compared them with the published Hi-C map of the model alga *Ectocarpus*[31] (**Table S1**). In a first step, we focused on the Hi-C maps from male individuals to perform the comparison across species. We then included male and female samples from dioicous species to examine potential sex differences in the 3D genome, together with derived monoicous (hermaphrodite) species to study the consequences of transitions to monoicy (and therefore loss of sex chromosomes) (**Table S2-S4**).

Biological replicates of Hi-C experiments were generated for each species and sex (**Fig. S4**). In total, we obtained between 263.5 and 892.5 million Hi-C interaction read pairs, depending on genome size and library efficiency (**Table S4**). Quality assessments showed strong reproducibility between biological replicates, with consistent cis-interaction frequency patterns (**Fig. S5A**) and Stratum-Adjusted Correlation Coefficient (SCC) scores[42] (**Fig. S5B**). Distance-dependent contact frequencies (P(s)) and slope of P(s)[43] reveal similar decay patterns for both replicates in all cases (**Fig. S5C, S5D**). Given the high reproducibility observed among biological replicates, the datasets were merged to maximize sequencing depth, thereby enabling downstream comparative analyses at higher resolution (**Fig. S6)**.

We examined the global chromatin compartmentalization across species using Hi-C contact maps. In the 3D nuclear space, genomes are typically organized into A (active) and B (inactive) chromatin compartments[3]. We derived A/B compartment profiles for each species by performing eigenvector decomposition (**Fig. 2A, Fig. S7**) which showed that all species display compartmentalized chromosomes. Previous studies have shown that compartmentalization strength can vary across organisms and resolution scales[44]. Given the range of genome sizes among the species analyzed, we asked whether genome size influences compartment strength in brown algae. Compartment strength analysis using saddle plots (see Methods; **Fig. 2B**) revealed that the morphologically complex alga, *U. pinnatifida* (BB: 2.21, AA: 1.24; genome size: ~534 Mb) and *D. herbacea* (BB: 2.10, AA: 1.18; genome size: ~422 Mb), exhibited the strongest compartmentalization. In contrast, *S. ischiensis*, the morphologically simple outgroup species, showed the weakest compartmentalization (BB: 1.45, AA: 1.07; genome size: ~195 Mb). *Ectocarpus* sp. 7 (BB: 1.62, AA: 1.56; genome size: ~186 Mb) and *D. dichotoma* (BB: 1.47, AA: 1.01; genome size: ~848 Mb) displayed comparable, intermediate compartment strengths despite their markedly different genome sizes. This contrast indicates that compartmentalization strength does not scale with genome size (**Fig. 2C, D**). Together, these results suggest that compartment strength is primarily shaped by species-specific genome organization and may be linked to increased morphological complexity rather than genome size alone.

To assess whether differences in compartment strength between species reflect broader changes in compartmental organization, we analyzed the proportion of A and B compartments and the distribution of eigenvector values across species. Given that larger genomes typically harbor more

TEs[25,26], which are often associated with heterochromatin and B compartments, a higher prevalence of B compartments in species with larger genomes is expected. However, our analysis revealed strikingly similar A/B compartment fractions across all species examined, with no clear correlation between genome size and compartment proportions (**Fig. 2D**). Therefore, both the organization and relative abundance of A/B compartments are largely independent of genome size, even though their strength may vary. Note that consistent with the open/closed nature of A/B compartments, compartment A has more exonic sequence and compartment B has more repeats, and genes located within compartment A exhibited significantly higher expression levels than those in compartment B across all species analyzed (**Fig. 2E**).

Previous work reported that the *Ectocarpus* genome lacks small-scale chromatin structures such as TADs[31]. Unexpectedly, visual inspection of Hi-C maps revealed domain-like structures in *U. pinnatifida*, *D. herbacea* and *D. dichotoma* (**Fig. S8A**). These domains, which exhibit strong self-interaction frequencies and are delimited by pronounced insulation boundaries, ranged in size from 258.2 kb in *D. dudresnayi* to 355.1 kb in *U. pinnatifida* (**Fig. S9**). To further characterize these structures, we generated insulation profiles using multiple distance ranges (50-500 kbp, **Fig. S8B**). Based on visual inspection, cis-chromatin contacts within 100 kbp provided optimal boundary resolution and were used for subsequent analysis, leading to the identification of insulated regions that we named 'boundaries'. To assess whether genome size relates to chromatin organization, we compared the mean insulation strength of boundaries with genome size and observed a moderate positive correlation (Pearson's $r = 0.53$), although this association was not statistically significant ($p = 0.223$), likely due to the small sample size (**Fig. S8C**). No enrichment of genomic features, including TEs, exons, or introns, was observed at these boundaries (**Fig. S8D**). Therefore, although morphologically complex species with large genomes tended to exhibit domain-like structures, these could not be linked to any specific genomic feature.

In the outgroup species *S. ischiensis*, the Hi-C contact map revealed plaid patterns indicative of alternating open and closed chromatin regions, as well as centromere interaction clusters and centromere aggregations (**Fig. S6**). No self-interacting chromatin domains were detected, even at higher resolutions (**Fig. S8A**). However, we observed prominent inter-chromosomal interaction clusters involving *S. ischiensis* chromosome 5. These regions are enriched for LINE retrotransposons (see **Fig. S10**).

Brown algae exhibit a relatively conserved chromosome number, ranging from 27 to 32. Macrosynteny analyses reveal largely 1:1 orthology across most chromosomes when compared to outgroup species (**Fig. 1C**). Given this conservation, we asked whether it extends to 3D chromatin architecture. Hi-C matrices were visually inspected in Juicebox[45] to assess whether orthologous chromosomes could be distinguished based on contact patterns. Notably, *D. herbacea* and *D. dudresnayi*, which diverged ~11 Mya, display clear 1:1 synteny, composed of a few large syntenic blocks (see **Fig. 1C**), and their orthologous chromosomes share remarkably similar inter-chromosomal interactions (**Fig. S6F-G**). To quantify this similarity at the 3D chromatin level, Hi-C contact matrices were normalized to uniform 50 kb bins. Pearson correlation coefficients (PCCs)

across the 32 orthologous chromosomes ranged from 0.012 to 0.53 (**Fig. 2F**), with higher PCCs reflecting substantial conservation of chromatin organization (**Fig. 2G**). By contrast, comparisons between other pairs of sister species such as *Ectocarpus* and *C. linearis* showed lower PCCs (**Table S6**), likely due to greater divergence times (110 My). These results indicate that chromatin conformation is largely conserved for at least ~11 million years of brown algal evolution, but not over larger evolutionary timescales.

In summary, A/B compartmentalization and its functional associations with gene density, repeats, and gene expression are largely conserved across 450 My of evolution encompassing brown algae and the outgroup *S. ischiensis*. By contrast, domain-like structures vary: species with greater morphological complexity show stronger, more distinct compartmentalization, whereas simpler species, including the outgroup, exhibit weaker domains.

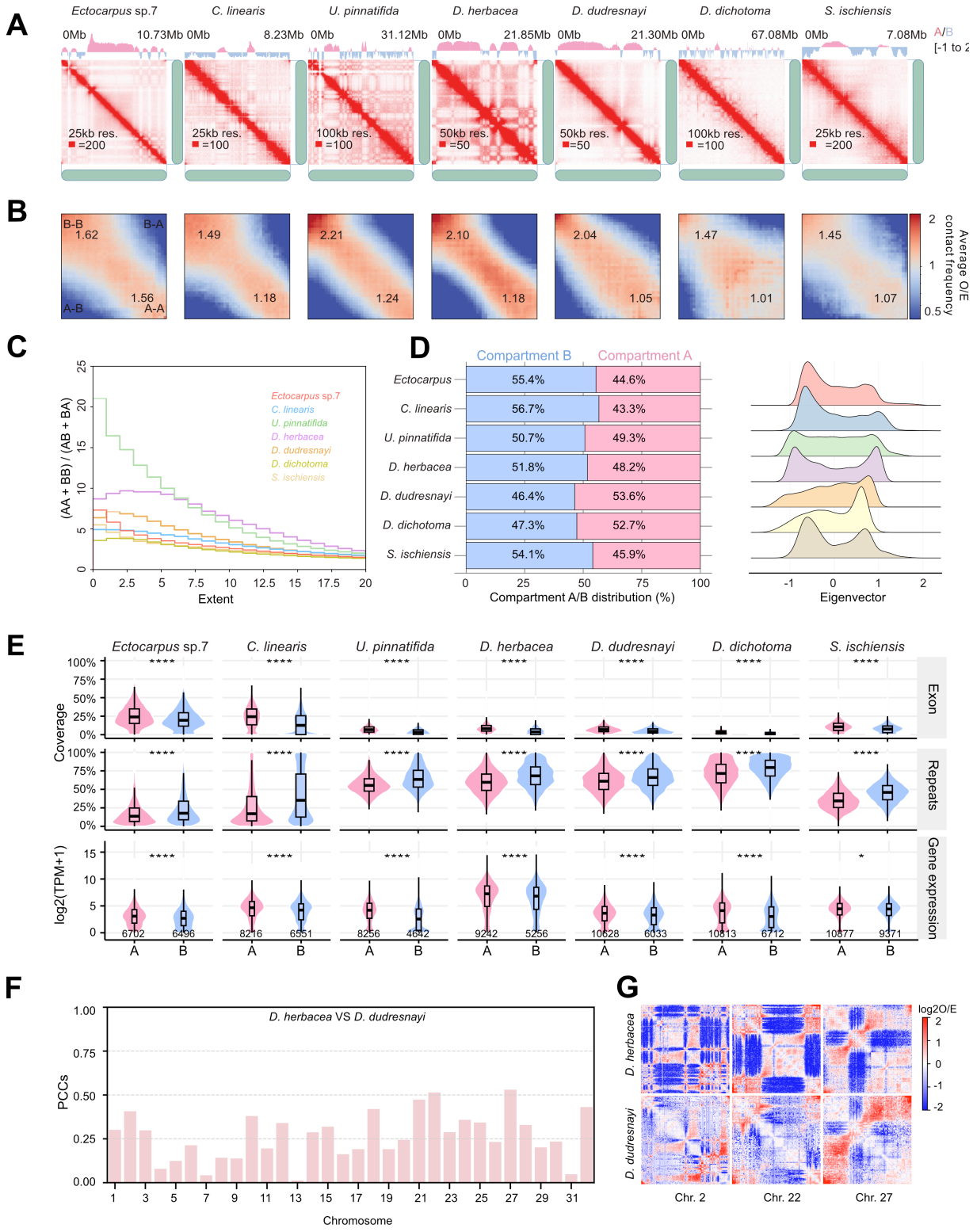



**Fig. 2. Comparison of chromosome conformation in brown algae**. (**A**) Chromosome 1 of each brown algal species with genome compartmentalization. Eigenvector values for each chromosome are shown above the corresponding Hi-C matrix in juicebox[45]. (**B**) Saddle plots of genome-wide interactions showing compartment contact interactions in cis for each species (numbers are relative interactions for AA and BB interactions type). (**C**) Saddle strength quantifying

compartmentalization interaction frequencies. The metric represents the ratio of intra-compartment (A–A and B–B) to inter-compartment (AB or BA) interactions, normalized by expected contact frequencies. Genomic bins are sorted according to eigenvector values obtained from Hi-C eigenvector analysis. The 'extent' parameter defines the range of eigenvector quantiles included in the calculation: an extent of 0 considers only the most extreme A and B bins, thereby capturing the strongest compartmentalization signal, whereas increasing extent values progressively incorporate bins closer to the center of the eigenvector distribution, where compartment identity is weaker. (**D**) Genomic distribution of chromatin compartments. Bars show the proportion of each genome assigned to A compartments (gene-rich, transcriptionally active) or B compartments (gene-poor, heterochromatic) based on Hi-C eigenvector analysis. (**E**) Violin plot shows the percentage of exons, repeats, and gene expression in each compartment across species. Numbers below brackets represent the number of genes. The lower and upper hinges of the box correspond to the first and third quartiles (the 25th and 75th percentiles). The upper whisker extends from the hinge to the largest and smallest values no further than 1.5x IQR from the hinge (Inter-Quartile Range, distance between the first and third quartiles). Significance was determined by a two-sample Wilcoxon rank sum test (****: $p < 0.0001$, *: $p <= 0.05$). (**F**) Pearson coefficient correlations (PCCs) of linearly transformed Hi-C contact matrices between homologous chromosomes of *D. herbacea* and *D. dudresnayi* across all chromosomes; (**G**) Examples show strongly correlated orthologous chromosomes between *D. herbacea* and *D. dudresnayi* at 50k resolution.

## Unique chromatin interaction patterns of sex chromosomes or sex-homologs

Like many early-diverged plants[23,46], sex determination in most brown algae occurs in the haploid stage of the life cycle, and haploid individuals of dioicous species possess either female (U) or male (V) sex chromosomes[23,29]. Previous analyses have shown that U/V sex chromosomes exhibit unusual linear genomic features, including distinct repeat content, gene density, and levels of sequence divergence compared with autosomes[25,47]. To explore whether these distinctive linear properties are reflected in higher-order chromatin organization, we analyzed the 3D topology of the U and V chromosomes across our sampled species. We examined contact frequencies, domain formation, and compartmentalization patterns, aiming to determine whether sex chromosomes adopt unique spatial conformations that could relate to their specialized genomic architecture and evolutionary dynamics.

Intriguingly, we observed prominent inter-chromosomal interactions between the sex-determining region (SDR) and multiple autosomes in *D. dichotoma*, *D. herbacea*, and *D. dudresnayi* (**Fig. 3A, Figs. S6, S11A**), suggesting a lineage-specific propensity for the sex chromosome to engage in trans-contacts, and this is not constrained by chromosome size (**Fig. S12A-B**). Note that these interactions do not result from self-interactions of tandem repeats (**Fig. S11B**) or mapping artifacts (**Fig. S11E**), but instead represent *bona fide* preferential contacts. To comprehensively investigate whether interacting regions share genomic features, we profiled the distribution of TE families, exons, compartment A/B regions, and centromeric sequences genome-wide for *D. dichotoma* (**Fig. S11F**). This analysis indicated that chromatin interactions are not primarily driven by these specific genomic features, indicating that alternative organizational principles may govern 3D genome structure in sex chromosomes.

Sex chromosomes were predominantly assigned to the B compartment across species (**Fig. S13A-C**), while sex-determining regions (SDRs) show enrichment in compartment A. Unexpectedly, repeat elements are more enriched in compartment A than compartment B on sex chromosomes,

inverting the typical genome-wide pattern where compartment A is gene-rich and compartment B is repeat-rich (**Fig. S13D**). Gene expression patterns in A versus B compartments vary across species, with significantly higher expression in A compartment observed only in *Ectocarpus* (**Fig. S13E**).

All examined SDRs exhibit strong insulation, indicating this is a conserved feature of brown algal sex chromosome organization (**Fig. 3B**). Notably, *D. herbacea* shows sex-specific differences in chromatin architecture: the female SDR is highly organized into smaller, self-interacting chromatin domains, while the male SDR lacks this organization (**Fig. 3B**). Genes located at chromatin domain boundaries displayed elevated expression compared to those within or outside domains on sex chromosomes (**Fig. S14**). Together, these data demonstrate that brown algal sex chromosomes have evolved conserved yet specialized chromatin architectures featuring enhanced insulation and reorganized domain structures, establishing 3D genome organization as a fundamental mechanism underlying sex chromosome evolution and regulation.

We next asked whether sex chromosome chromatin architectures persist following evolutionary transitions to hermaphroditism, where former sex chromosomes become autosomes[48]. Our dataset captures two independent transitions: *Ectocarpus* versus *C. linearis*, and *D. herbacea* versus *D. dudresnayi*, with the latter representing a recent transition occurring within the last 11 million years. Strikingly, in both lineages, the chromatin organization of former sex chromosomes ("sex-homologs") has retained the architectural signature of the ancestral male (but not female) sex chromosome rather than adopting autosomal patterns (**Fig. S15**). This evolutionary constraint is most pronounced in *D. herbacea*, where the large SDR and recent transition timeline provide a clear window into chromatin evolution (**Fig. 3B**). These findings demonstrate that three-dimensional chromatin organization exhibits remarkable evolutionary inertia, with former sex chromosomes maintaining male-specific architectural features long after their functional role in sex determination has been lost.

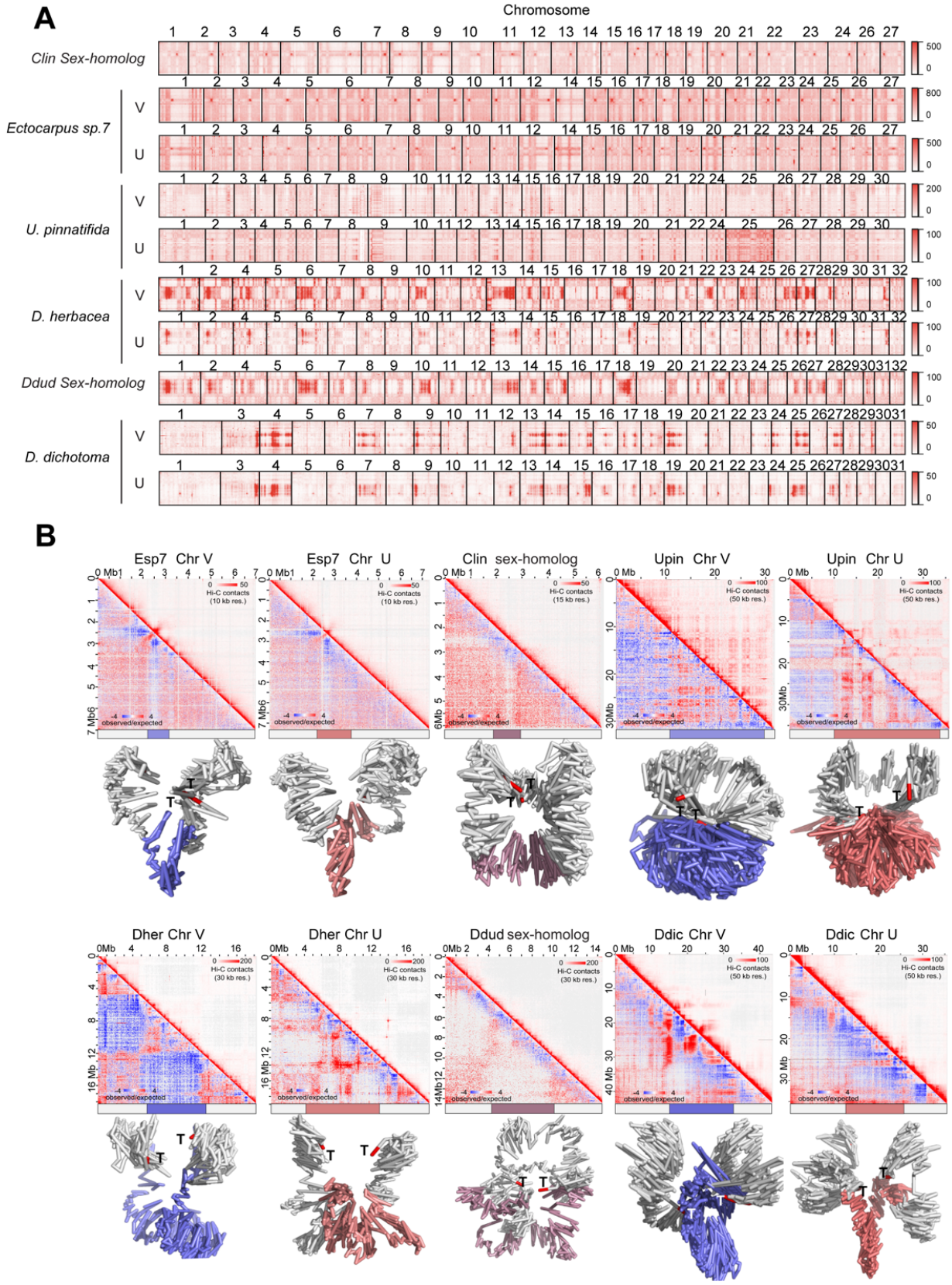



**Fig. 3.** 3D chromatin configuration of sex chromosome or sex-homologs in brown algae. (A) Hi-C contact maps

showing inter-chromosomal interactions between the sex chromosome (or sex homolog) and autosomes for each species; (B) Hi-C map and reconstructed 3D configurations of sex chromosomes at 25 k resolution Hi-C data, maximum likelihood approach was employed to construct 3D structures from Hi-C data of *Ectocarpus* sp.7[31], *C. linearis*, *U. pinnatifida*, *D. herbacea*, *D. dudresnayi* and *D. dichotoma* with a default setting in 3DMax[49]. SDRs or SDR-homologs in the V, U chromosomes and sex-homolog are colored in slate, deepsalmon and purple, respectively, and telomeres are colored in red, pseudoautosomal regions (PARs) are colored in grey.

## Longterm co-evolution of centrophilic retrotransposons and brown algal centromeres

The structures of regional centromeres exhibit extreme variation among eukaryotes and typically evolve rapidly[50,51]. The centromeres of many species feature centrophilic retrotransposons, specific TE families that target the centromere and in some cases directly correspond to the sequence occupied by the centromeric histone variant CENH3[52]. Centrophilic retrotransposons can co-occur with centromeric satellite repeats, forming a spectrum from a relatively minor contribution to centromeric sequence, e.g., *ATHILA* elements in *Arabidopsis thaliana*[53], to dominating the centromeres of many chromosomes, e.g. the *CRM* elements of maize[54]. In other species, centrophilic retrotransposons solely define and constitute the centromeres, e.g. *Bryco* elements in the moss *Physcomitrium patens*[55] or *ZeppL* elements in the green alga *Chlamydomonas reinhardtii*[56,57].

We previously identified centromeric regions in *Ectocarpus* sp. 7 based on strong centromere-centromere interactions in inter-chromosomal Hi-C contact maps[31]. Although these interaction clusters span hundreds of kilobases, we mapped the putative centromeres to highly localized regions that typically lack satellite DNA but are defined by the presence of two specific LTR retrotransposon families from the *Metaviridae*/Ty3 group, termed *ECR* (*Ectocarpus Centromeric Retrotransposon*) elements. **Fig. 4A** shows a representative pair of chromosomes from *Ectocarpus* sp. 7, with a single broad region of strong centromere-centromere interaction clusters visible on inter-chromosomal Hi-C matrix in Juicebox[45], in addition to prominent telomere-telomere interaction clusters. The two centromeric regions harbor short (~30-40 kb) clusters of the *ECR-1* (orange) and *ECR-2* (vermillion) elements, and with no other repeats consistently present these retrotransposons presumably correspond to the epigenetic centromere (i.e., the region featuring CENH3). While many *ECR-1* elements are intact and potentially functional, all *ECR-2* copies are degraded and the family may be in the process of extinction[31].

We observed prominent centromere-centromere and telomere-telomere interaction clusters in the Hi-C contact maps of all species (**Fig. 4B-E, S6**). This organization appears to be common to both brown algae and the outgroup *S. ischiensis*, and genome folding on a broad chromosomal level may be consistent across the whole lineage. We searched for homologs of *ECR-1* in each genome and asked whether these elements also form clusters within the broad centromeric regions defined from the Hi-C maps. In *U. pinnatifida*, we identified two homologous families: *ECR-1_uPin* mostly (but not exclusively) forms short centromere-localized clusters (**Fig. 4B and S16C,** orange), whereas *ECRl-2_uPin* exhibits a genome-wide distribution (**Fig. 4B and S16C,** blue). In *D. herbacea*, the families *ECR-1_dHer* and *ECR-2_dHer* both co-occur in short centromeric clusters (**Fig. 4C, S16D**). Although Hi-C contact maps are unavailable, we also identified *ECR* families

that either form highly localized clusters or have genome-wide distributions in the contiguous genome assemblies of the species *Scytosiphon promiscuus* and *Fucus serratus*[25] (**Fig. S16B and S16F**), mirroring the pattern observed in *U. pinnatifida*. In *C. linearis* the situation is less clear; we identified a single *ECR* family that is localized within the Hi-C defined centromeric regions (**Fig. S16**A, **S6A**), however it is only present on 10 of the chromosomes and is frequently fragmented, suggesting that in may be in the process of elimination from the genome. At least two other repeats are enriched within the broadly defined centromeric regions, although these resemble neither TEs nor satellites and their role (if any) in centromere function is unclear. One appears to be a multicopy gene encoding a protein with a predicted P-loop domain (Pfam PF07999).

Conversely, in *D. dichotoma* we identified three abundant *ECR* families that all exhibit genome-wide distributions (**Fig. 4D and S16E**). Similarly, in *S. ischiensis* the sole *ECR* family does not exhibit centromeric clustering and is present at low copy numbers elsewhere in the genome (**Fig. 4E and S16G**).

We performed a phylogenetic analysis using the combined Gag and Pol protein sequences of all identified *ECR* elements. Interestingly, the centromere-associated *ECR* elements form a distinct and robustly supported sub-lineage (ultrafast bootstrap value 100) within the wider diversity of *ECR* elements (**Fig. 4F**). Thus, the lineage can be divided to genuine centrophilic *ECR* elements and non-centrophilic *ECR-like* elements (**Fig. 4F**), which can co-occur in a single genome (*U. pinnatifida, S. promiscuus* and *F. serratus*), be present only as centrophilic *ECR* elements (*Ectocarpus* sp. 7 and *D. herbacea*), or only as *ECR-like* elements (*D. dichotoma* and *S. ischiensis*). Thus, it appears that centrophily evolved from a non-centrophilic LTR retrotransposon and had emerged in the common ancestor of the brown algal crown radiation (BACR) at least 160 MY[22].

We next searched for any repeats that are enriched in the broadly defined centromeric regions of *D. dichotoma* and *S. ischiensis*. Surprisingly, we identified single families of Copia (*Pseudoviridate*/Ty1) LTR retrotransposons that form discrete clusters in the two species (**Fig. 4D, E**). Phylogenetic analysis of all Copia elements identified in the six analyzed species revealed that these putatively centrophilic elements do belong to the same weakly supported lineage (**Fig. S16H**). However, this lineage also includes three other Copia family from *S. ischiensis* that have a genome-wide distribution, and families from *Ectocarpus* sp. 7, *D. herbaceae* and *C. linearis* that are present genome-wide. Thus, it is possible that there have been two independent transitions to centrophily during the evolution of *D. dichotoma* and *S. ischiensis*, or alternatively, the centrophilic Copia elements could represent an ancestral state with multiple transitions to centrophobic insertion patterns.

Despite featuring distinct LTR retrotransposons, there are parallels between the putative centromere structures of most brown algae and *S. ischiensis*. The LTR clusters are generally short, typically spanning tens of kilobases as opposed to hundreds of kilobases as common in other species[58]. This appears to be true regardless of the repeat content of the genome, for example, in the more repeat-rich genomes of *U. pinnatifida, D. herbacea* and *D. dichotoma*, the centromeres do not exhibit an elevated repeat content relative to the surrounding sequence (**Fig. 4B-D**). As described

for *Ectocarpus* sp. 7[31], satellite DNA is also conspicuously absent from most centromeres in all species (**Fig. 4A-E**), reinforcing the interpretation that the centrophilic retrotransposons (either *ECR* or Copia) may directly correspond to the epigenetic centromeres across brown algae.

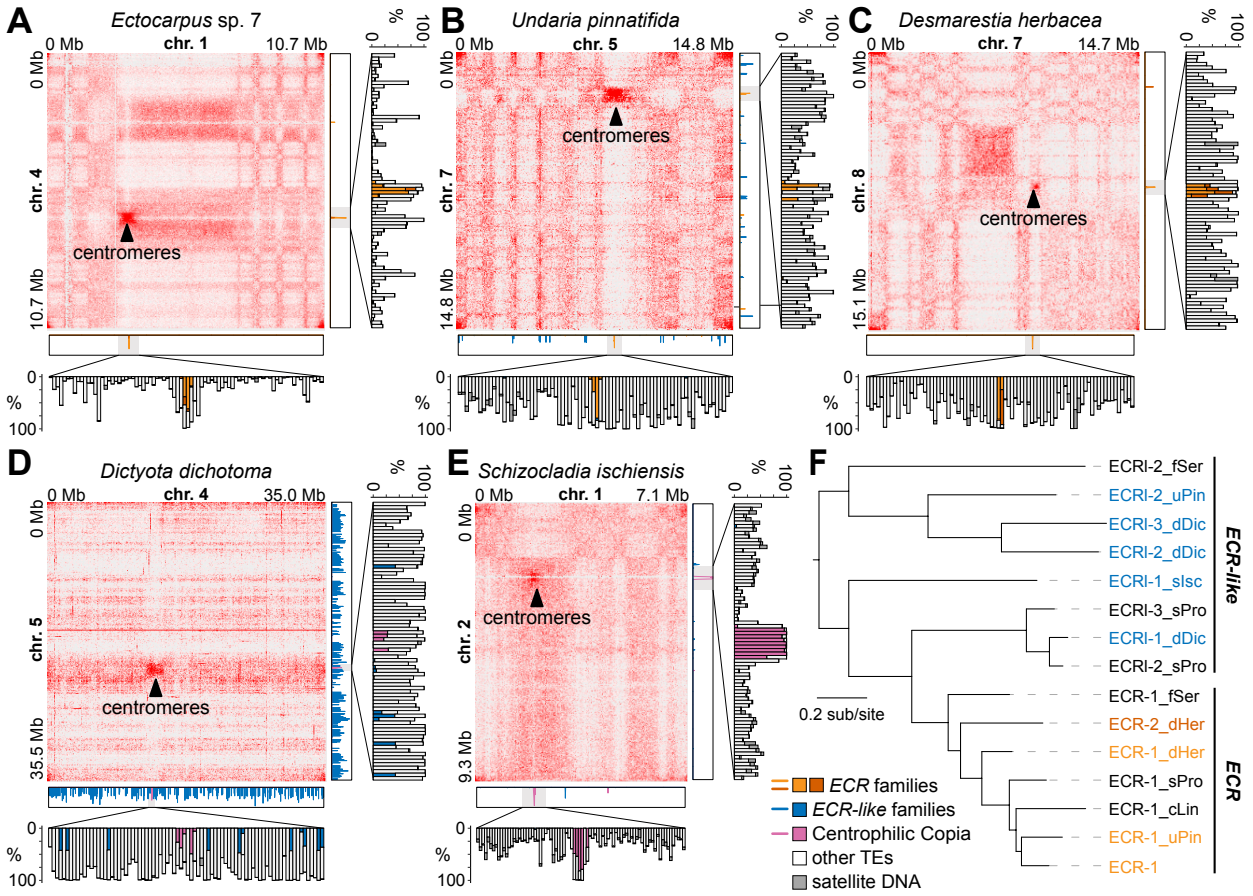



**Fig. 4.** Centromere profiles across brown algal genomes. Representative examples show centromere clustering between two chromosomes in Hi-C interaction frequency maps, along with the distribution of centromeric retrotransposons (*ECR* and Copia), non-centrophilic *ECR-like* elements, other TEs and satellite DNA in 10 kb windows for (A) *Ectocarpus* sp. 7, (B) *U. pinnatifida*, (C) *D. herbacea*, (D) *D. dichotoma*, and (E) *S. ischiensis*. Chromosome-scale plot shows only centromeric retrotransposons and *ECR*-like elements, centromere-centered panel shows all repeat types. (F) Maximum likelihood phylogeny of *ECR* and *ECR-like* elements constructed from Gag and Pol protein sequences under the LG+F+I+G4 model. All *ECR* elements form genomic clusters, and those forming specific centromeric clusters in panels A-E are colored orange or vermillion. All nodes received > 95% ultrafast bootstrap support with the exception of those connecting *ECR-1*, *ECR-1_uPin* and *ECR-1_cLin*.

## The ancestral brown algal karyotype and accelerated chromosome evolution in the Ectocarpales

Chromosome evolution involves major genome rearrangements that occur within (e.g., inversions) or between chromosomes, with the latter category including several types of chromosome fusion and fission, in addition to reciprocal translocations. As introduced above, chromosome number ranges from 27 to 32 in our sample of genomes, and the availability of chromosome-level assem-

blies enables us to reconstruct chromosome evolution over ~240 million years of brown algal evolution for the first time. As presented in **Fig. 1C**, one-to-one chromosome-scale synteny is evident among the majority of chromosomes (n=21) in a comparison of *D. dichotoma, D. herbacea* and *D. dudresnayi*, and *U. pinnatifida*. This number increases to 26 chromosomes in a pairwise comparison of *D. dichotoma* (here the outgroup) and *D. herbacea*, which appear to be differentiated by only three inter-chromosomal rearrangements. Although the order of these events cannot be determined without additional outgroups, if we arbitrarily assume that *D. dichotoma* represents the ancestral state then the rearrangements can be expressed as: i) fission of *D. dichotoma* chromosome 6, giving rise to *D. herbacea* chromosomes 29 and 30, ii) fission of *D. dichotoma* chromosome 1 and subsequent fission of one fragment to chromosome 12, giving rise to *D. herbacea* chromosome 30, and iii) reciprocal translocation between *D. dichotoma* chromosomes 27 and 31, yielding *D. herbacea* chromosomes 21 and 24 (see **Fig. 1C**). Furthermore, based on parsimony we can infer that the ancestral brown alga had either 31 or 32 chromosomes, with the point of distinction our inability to determine if *D. dichotoma* chromosome 6 is the result of a fusion, or if *D. herbacea* chromosomes 29 and 30 are the product of a fission.

The remaining inter-chromosomal rearrangements can be polarized relative to this ancestral karyotype. *D. dudresneyi* features one unique rearrangement event, the fission of *D. herbacea* chromosome 30 and subsequent fusion of a fragment to chromosome 19, which can be inferred to have occurred in the last 11 million years. As with the inferred fission+fusion event between *D. herbacea* and *D. dichotoma*, it is also possible that this event was instead a reciprocal translocation that occurred close to the telomere on one of the chromosomes, leaving no trace in the synteny analysis. The *U. pinnatifida* chromosomes 1 and 25 are both products of fusion events that are unique to this species, whereas chromosomes 12 and 29 result from a reciprocal translocation that is also shared by *Ectocarpus* sp. 7 and *C. linearis*, suggesting it occurred prior to the common ancestor of Ectocarpales and Laminariales. Strikingly, the Ectocarpales feature the most derived karyotypes, especially *C. linearis* (**Fig. 1C**). Although the complexity of several of these rearrangements complicates inference, we estimate that two fusions occurred in the ancestor of Ectocarpales, with two further fusions and a reciprocal translocation occurring on the lineage leading to *Ectocarpus* sp. 7, and two distinct fusions and seven reciprocal translocations occurring on the lineage leading to *C. linearis* (**Fig. 5A, S17**). Despite the higher rate of rearrangements in the Ectocarpales, 8 chromosomes exhibit a 1-1 relationship across all analyzed brown algal genomes, including the sex (or sex-homolog) chromosome.

Centromeres are frequently involved in chromosomal rearrangements, with centromere-proximal rearrangement breakpoints associated with end-to-end fusions (including Robertsonian translocations) and nested fusions (i.e., centric insertion)[17,59]. To further address the origin of brown algal chromosomal rearrangements, we mapped centromere locations on to a representative selection of rearrangements (**Fig. 5A-C**). To determine if centromere locations are conserved among species, we performed an ancestral genome reconstruction using AGORA, which assembles contiguous ancestral regions (CARs) featuring multiple genes inferred to have been adjacent in an ancestral genome[60]. This analysis revealed that centromeres are generally located within the same CAR, or

between the same adjacent CARs, in each genome, suggested ancient conservation of centromeric locations across brown algae (**Fig. 5A-C**). This is true even for *D. dichotoma*, despite the difference in the identity of centrophilic LTR retrotransposons in this species (see **Fig. 4D**).

Most reciprocal translocations were not associated with centromeres, as exemplified by the reciprocal translocation common to *U. pinnatifida* and *Ectocarpus* sp. 7 (**Fig. 5A, B**). Conversely, many of the fusions (or fissions) potentially involved centromeres. The *D. herbacea* chromosomes 29 and 31 are both acrocentric, suggesting that the fusion or fission event that differentiates them from *D. dichotoma* could have been a Robertsonian translocation (**Fig. 5C**). However, at least one centromeric inversion has likely occurred since this event, complicating inference. *U. pinnatifida* chromosome 1 appears to have resulted from centric insertion of the acrocentric *D. herbacea* chromosome 31 into the centromere of *D. herbacea* chromosome 15 (**Fig. 5C**). Similarly, *U. pinnatifida* chromosome 25 is the product of a nested insertion of *D. herbacea* chromosome 19 into 9, although in this case the insertion point appears to have been adjacent to the centromere rather than within it, with the ancestral chromosome 19 centromere having been lost in the fused chromosome (**Fig. 5D**). Note that *Ectocarpus* sp. 7 chromosome 7 is the product of an end-to-end fusion between the acrocentric *D. herbacea* chromosome 31 and the metacentric *U. pinnatifida* chromosome 29, which resulted in the loss of the chromosome 31 centromere (**Fig. 5C, D**). Many of the other rearrangements in Ectocarpales are also adjacent to a centromere on at least one of the chromosomes (**Fig. S17**).

Finally, we used the ancestral brown algal genome reconstruction to investigate syntenic relationships with *S. ischiensis*, which last shared a common ancestor with brown algae ~450 mya (see **Fig. 1B**). Arbitrarily assuming that *D. herbacea* represents the ancestral state of brown algal chromosomes, we painted the *S. ischiensis* chromosomes based on the orthology relationships of genes between the two species. Essentially all of the *S. ischiensis* chromosomes feature a mixture of genes from different brown algal ancestral chromosomes (**Fig. S18**), with the possible exceptions of chromosomes 24 and 26 that correspond to *D. herbacea* chromosomes 17 and 26, respectively, with only limited mixing with other chromosomes. This analysis shows that *S. ischiensis* shares low macrosynteny with brown algal genomes, confirming a result derived previously from a more fragmented version of the genome assembly[25].

Overall, our results reveal that the ancestral brown alga likely had 31 or 32 chromosomes, with a small number of centromere-associated fusion/fissions and mostly non-centromeric reciprocal translocations differentiating the karyotypes of most species. The rate of chromosome evolution appears to have accelerated in the Ectocarpales, and especially in *C. linearis*, which exhibits the most derived brown algal karyotype resulting from extensive reciprocal translocation in this lineage (**Fig. 5A**).

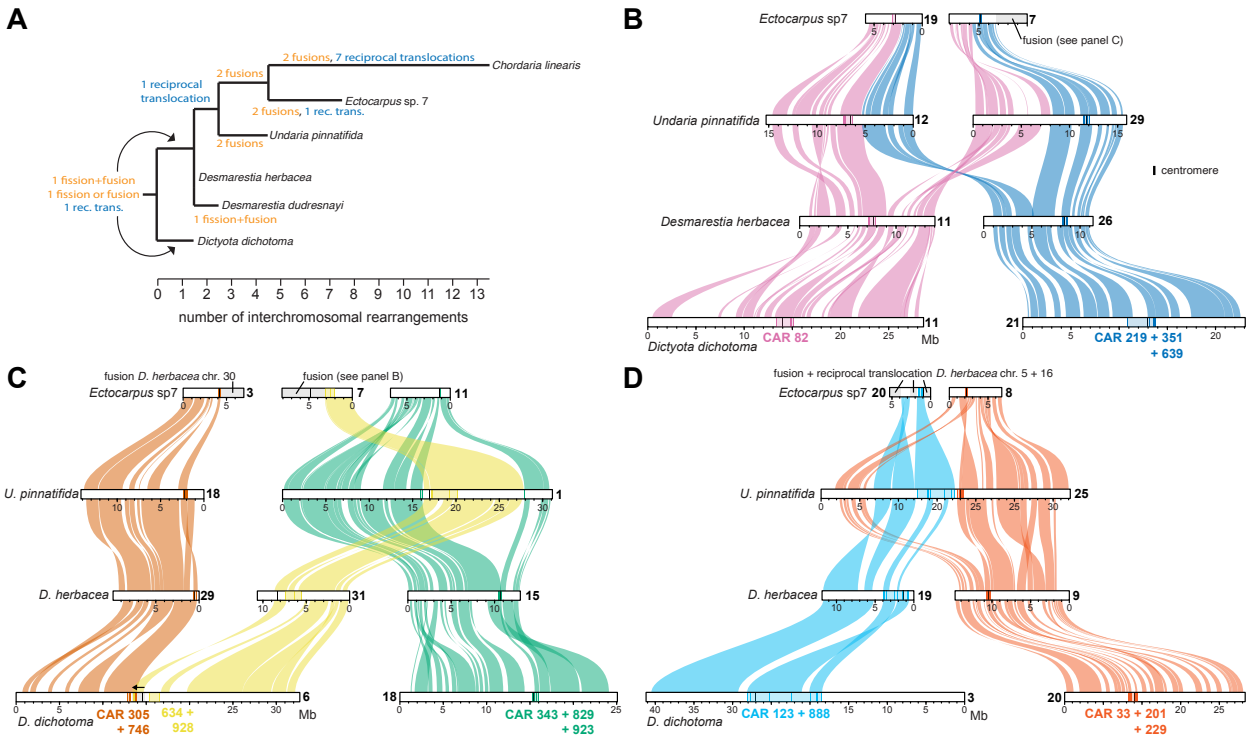



**Fig. 5**. **Inter-chromosomal rearrangements in brown algal genomes.** A) Phylogeny of analyzed brown algal genomes where branch lengths correspond to the number of inter-chromosomal rearrangements that occurred between each node. B-D) Syntenic relationships among a representative selection of chromosomes associated with inter-chromosomal rearrangements, showing a reciprocal translocation (B), a putative Robertsonian translocation (C) and nested fusions (C, D). Chromosomes are colored relative to *D. herbacea* and contiguous ancestral regions (CARs) defined by AGORA[49] that flank centromeres are highlighted (each line corresponds to a gene present in the CAR). A predicted centromeric inversion on *D. dichotoma* chromosome 6 (panel C) is represented by a black arrow. For all Ectocarpales rearrangements see **Fig. S17**.

# Discussion

## Conservation and variation in 3D chromatin architecture across evolutionary scales

Our analyses reveal a striking conservation of higher-order chromatin architecture despite wide variation in genome size and repetitive element content. These results suggest that key features of 3D genome folding: such as chromosomal territories, centromere clustering, telomere clustering, and A/B compartments are maintained under strong evolutionary constraint across the brown algal lineage, mirroring patterns in animals[3], plants[61], and fungi[62], and reflecting a conserved functional architecture for gene regulation and chromatin organization. The presence of A/B compartmentalization in brown algae, which diverged from plants and animals over a billion years ago[1], indicates either deep ancestral origins or convergent evolution driven of this feature by fundamental nuclear constraints. Critically, despite substantial variation in brown algal genome sizes neither eigenvector distributions nor compartmentalization strength correlate with genome size. This independence from genome composition suggests that spatial organization may serve essential regulatory functions that go beyond genomic complexity.

In animal lineages, chromatin domains are defined by local insulation boundaries and shaped by architectural proteins such as CTCF, cohesin, and condensin[1]. In contrast, both plants[36,61] and brown algae lack canonical insulator proteins like CTCF, and there are no prominent TADs as commonly observed in animals[63]. In plants with large genomes, however, domains can be found and their boundaries are often associated with transcriptionally active regions specific histone modifications[10,64] (e.g., H3K4me3 or H3K9me2) or chromatin state transitions, rather than fixed sequence motifs. Similarly, in brown algae, non-canonical chromatin domains are detectable in some species with larger genomes, exhibiting elevated insulation levels that suggest alternative, genome size or TE content-dependent mechanisms for domain formation. Notably, the relationship between gene expression and chromatin domains appears inconsistent across brown algal species, unlike in animals where genes near domain boundaries often exhibit elevated expression[1]. We found no evidence for enrichment of specific genomic features such as gene families or TE classes at insulation boundaries. This contrasts with other eukaryotic systems, where boundary elements often coincide with active genes, tRNAs, or specific DNA-binding motifs[34,65], though CTCF involvement varies and boundary mechanisms can differ among lineages. In brown algae, boundary positioning may instead be influenced by other, as yet unidentified factors, such as nucleosome organization[66], histone modifications[1], or structural RNAs[67]. Alternatively, boundaries may form passively as a consequence of chromatin state transitions or replication timing domains, as proposed in some plant species[62,68].

Taken together, our results demonstrate that A/B compartmentalization is a conserved and robust feature of brown algal nuclear organization, maintained over evolution despite substantial differences in genome size and repeat content. In contrast, chromatin domains defined by local insulation minima are observed only in larger genome species, suggesting that domain-level genome folding is not a universal characteristic within this lineage. The lack of correlation between genome size and insulation strength contrasts with models in other eukaryotes where genome expansion

and TE load are major drivers of architectural complexity. Our findings thus support the concept that 3D genome architecture is shaped by lineage-specific evolutionary constraints[44] rather than being universally determined by genome size alone.

**Conserved macrosynteny and chromosomal stability over evolutionary time**

We generated high-quality, chromosome-level genome assemblies for five brown algal species and one outgroup. These assemblies markedly improve completeness and contiguity over previous versions, enabling detailed analyses of genome evolution and 3D chromatin organization in this understudied eukaryotic lineage.

Ancestral reconstruction of chromosomes or linkage groups has emerged as a powerful analytical technique in many eukaryotic groups, revealing various degrees of conserved macrosynteny among mammals[27], vertebrates[69], and even some[17], but far from all[70], bilaterians. The concepts of macrosynteny and ancestral linkage groups have long been associated with the Muller elements of drosophilids, the gene content of which remains largely conserved over more than 50 million years of evolution despite the gene order itself being frequently rearranged between species by inversions[71]. Similar ancestral linkage groups have been defined for lepidopterans (Merian elements)[72] and rhabditid nematodes (Nigon elements)[73], as well as in some plant groups such as the Brassicaceae[74].

Our chromosome-level assemblies revealed comparably extensive macrosynteny across brown algae, enabling an ancestral karyotype of either 31 or 32 linkage groups to be constructed. These ancestral chromosomes had become established at least 240 million years ago, with earlier evolutionary relationships made uncertain by the extensive breakdown of macrosynteny between brown algae and the outgroup *S. ischiensis*. Gene collinearity (i.e., synteny) is also extensive among brown algal genomes (Fig. 1C, 5B-D), although some intrachromosomal rearrangements between species have occurred as in other eukaryotic groups with conserved macrosynteny. Notably, chromosomes that have arisen via fusion generally retain their ancestral macrosynteny patterns (Fig. 5C, D), with limited intrachromosomal rearrangements between the different ancestral linkage groups. While these may simply represent very recent events, similar patterns are observed in the Muller and Merian elements of insects[71,72], and it may be the case that macrosynteny is selectively retained to maintain *cis*-regulation between genes[27,72]. Overall, as in other complex eukaryotic lineages, conserved and ancient macrosynteny may reflect strong evolutionary constraints contributing to long-term genomic and adaptive stability.

Nevertheless, we detected substantial chromosomal rearrangements in the Ectocarpales, corresponding to elevated rates of both fusions and reciprocal translocations. Interestingly, this group comprises species with the smallest and most compact genomes. Such compact genomes, typically characterized by higher gene density, are often linked to elevated recombination rates[75] which can in turn promote chromosomal rearrangements. Smaller chromosomes have also been more substantially involved in fusions in vertebrate[76] and lepidopteran[72] evolution. Moreover, Ectocarpales species have short life cycles[21] with frequent meiotic divisions, a feature that may further enhance recombination activity and thus accelerate the rate of chromosomal reorganization[77] observed in

this lineage. Finally, the Ectocarpales also exhibit reduced developmental complexity relative to the other analyzed brown algae and although speculative we cannot rule out that this evolutionary transition to simpler morphologies has resulted in relaxed constraint on macrosynteny.

**Distinct chromatin features of sex chromosomes and sex-homologs**

Little is known about the 3D chromatin organization of sex chromosomes across eukaryotes[20,78]. Sex chromosomes in brown algae exhibit distinct three-dimensional architectures that set them apart from autosomes. Although largely assigned to the transcriptionally inactive B compartment, their non-recombining SDRs consistently localize to the active A compartment. This organization is further marked by an inversion of genomic feature distribution, with repeat elements enriched in active A compartments. It is possible that this reflects the conspicuous presence of TEs in SDR gene introns[25,47,40]. Strong insulation of all examined SDRs points to a conserved structural hallmark of non-recombining regions across lineages. Evolutionary analyses further reveal that former sex chromosomes retain male-, but not female-, specific chromatin signatures long after transitions to hermaphroditism, underscoring the persistence of ancestral features[25,79]. In brown algae, hermaphrodites are derived from male lineages that acquired female-specific genes[25,48]. Accordingly, our results suggest that V-specific chromatin architecture is maintained through this transition. Together, these results show that chromatin organization is both a defining feature of sex chromosomes and an evolutionarily constrained trait that endures beyond its original functional role.

**Conserved centromeric locations and long-term co-evolution with centrophilic retrotransposons**

The extensive centromere-centromere contacts in the Hi-C maps of brown algae and *S. ischiensis* enabled us to investigate centromere structure over 450 million years of evolution. As we recently established in *Ectocarpus* sp. 7[31], the centromeres of all species feature a short cluster of LTR retrotransposons that belong to specific families that are conspicuously absent from the rest of the genome. These centrophilic LTR families have presumably evolved to target the centromere of their respective genome, as has been established for several distinct groups of retrotransposons in specific plant and animal genomes[52,80]. The absence of satellite DNA or any other consistent repetitive sequences also suggests that the centrophilic retrotransposon clusters likely correspond to the epigenetic centromere, as in *P. patens*[44] and *C. reinhardtii*[57]. Antibodies targeting CENH3, which are currently unavailable for brown algae, will be required to test the association between the centrophilic LTR elements and the epigenetic centromere.

Our analyses also revealed the long-term co-evolution between *ECR* LTR retrotransposons and the centromeres of a subset of species that correspond to the brown algal crown radiation (BACR). *ECR-1*, the intact and presumably active centrophilic retrotransposon of *Ectocarpus* sp. 7[31], forms a distinct evolutionary lineage relative to all other *Metaviridae*/Ty3 families in the *Ectocarpus* sp. 7 genome[31]. We identified several other members of this lineage across the brown algal genomes, including both centrophilic and non-centrophilic families. Strikingly, the centrophilic elements, here referred to as genuine *ECR* elements, form a clade within the wider diversity of non-centrophilic *ECR-like* elements, suggesting an evolutionary transition to centrophily that occurred at least

160 mya in the common ancestor of the BACR. Although their long-term persistence in the genome of *C. linearis* is unclear, the centrophilic *ECR* elements appear to be the core centromeric components of all other BACR genomes analyzed, and searches for *ECR* elements are likely to prove useful for centromere identification across the majority of brown algal species.

Such long-term co-evolution between a specific lineage of retrotransposons and centromeres is most closely paralleled by the *CRM* elements that are present at the centromeres of many angiosperm species. *CRM* elements represent a centrophilic lineage within a wider *Metaviridae*/Ty3 retrotransposon clade known as the chromoviruses[81,82]. Furthermore, non-centrophilic relatives of centrophilic *CRM* elements are present in several genomes including those of gymnosperms[82], mirroring the presence of *ECR-like* elements in *U. pinnatifida*, *D. dichotoma* and *S. ischiensis*. Finally, chromoviruses typically encode a Pol protein that feature a chromodomain fused to the C-terminus of the integrase domain, which presumably targets integration to specific chromatin states[83]. The chromodomain has been replaced by a different putative targeting domain (the "CR" motif) in most *CRM* elements, although a role for this domain in centromeric targeting is yet to be experimentally established. We previously described a C-terminal chromodomain in the pol protein of *ECR-1*[31]. However, both the *ECR* and *ECR-like* elements recovered here encode an intact chromodomain, as do other independent non-centrophilic lineages of *Metaviridae*/Ty3 elements in brown algae[31]. Thus, it is unclear if the *ECR* chromodomain is responsible for centromere-targeting, and it is possible that the transition to centrophily evolved following amino acid substitutions in other regions of the integrase domain. Substitutions in the integrase domain between the centrophilic Copia LTR family *Tal1* and the closely related centrophobic family *Evade*, which do not encode chromodomains, was recently shown determine targeting specificity in *Arabidopsis* species[84].

Similarly, it is also currently unclear how centromere-targeting in the centrophilic Copia elements of *D. dichotoma* and *S. ischiensis* evolved. Phylogenetic analyses suggests that centrophily likely evolved independently in each species, since multiple related families with genome-wide distributions are extant in *S. ischiensis* and brown algae. Alternatively, we cannot rule out that the centrophilic Copia elements represent an ancestral state, with multiple reversions to genome-wide integration patterns occurring among related Copia families. Additional high-quality genomes from non-BACR species will be required to address this question.

Overall, our results capture several themes that are emerging from studies of centromere and retrotransposon co-evolution, including transitions in the identity of the major centrophilic families between species[52,84]. The evolution and ancient association of a chromodomain-encoding lineage of centrophilic *Metaviridae*/Ty3 elements represents a striking case of evolutionary convergence with the *CRM* elements of angiosperms. Spanning at least 160 million years, the ancient centrophily of *ECR* elements may only be exceeded in age by that of the *Bryco* Copia elements of moss centromeres[55].

Finally, although some centromeres have been lost following specific fusion events (especially in the Ectocarpales), we found that centromere locations were generally conserved among brown

algae. This is true even in *D. dichotoma*, suggesting that the transition between Copia and *Metaviridae*/Ty3 centrophilic retrotransposons did not alter centromere locations. Centromere conservation has received considerably less attention in ancestral genome reconstructions, although in some lineages this is impossible due to the presence of holocentric chromosomes. Centromere repositioning refers to the phenomenon of *de novo* centromere evolution without chromosomal rearrangement, which appears to occur relatively frequently in several mammalian and plant lineages[85–88]. Centromere locations have been conserved in the context of the Muller elements for more than 50 MY in *Drosophilia*, although *de novo* centromere evolution was recently shown to have occurred in the *ananassae* subgroup[89]. Similarly, ancestral Brassicaceae centromeres correspond to the extant centromeres in some species such as *Arabidopsis lyrata* and *Capsella rubella*[74], although some have been lost following fusions in species including *Arabidopsis thaliana*, and other brassica lineages are associated with extensive centromere repositioning[87]. However, few examples of centromere locations that have remained conserved for more than 200 million years have been reported, reinforcing the exceptional nature of macrosynteny in brown algal genomes.

## Methods

### Brown algae culture

Algae materials were cultured in autoclaved natural seawater (NSW) enriched with half-strength Provasoli nutrient solution (Provasoli-enriched seawater; PES) as previously described[30,90]. *C. linearis* (Clin) was grown at 14 °C with the light intensity of 25 µmol photons $m^{-2}$ $s^{-1}$ (8h light/16h dark); *U. pinnatifida* (Upin**)** male and female strains were grown at 14 °C with the light intensity of 25 µmol photons $m^{-2}$ $s^{-1}$ (12h light/12h dark); *D. herbacea* (Dher) male, female and *D. dudresnayi* (Ddud) strains were grown at 14 °C with the light intensity of 25 µmol photons $m^{-2}$ $s^{-1}$ (12h light/12h dark); *D. dichotoma* (Ddic) male and female strains were grown at 20 °C with the light intensity of 10 µmol photons $m^{-2}$ $s^{-1}$ (16h light/8h dark); *S. ischiensis* (Sisc) were grown at 20 °C with the light intensity of 25 µmol photons $m^{-2}$ $s^{-1}$ (16h light/8h dark). The medium was changed every two weeks.

### Hi-C library preparation

An *in situ* Hi-C protocol of plants was optimized for brown algae[31,61,91]. All strains of *C. linearis*, *U. pinnatifida*, *D. herbacea*, *D. dudresnayi*, *D. dichotoma*, *S. ischiensis* were cultivated in controlled lab conditions[25] and material was collected using a 40 µm filter and fixed in 2 % (vol/vol) formaldehyde for 30 min at room temperature, and the cross-linking reaction was quenched with 400 mM glycine. Approximately 50 mg of frozen algal tissue was ground in liquid nitrogen using a pre-chilled mortar and pestle. The resulting fine powder was suspended in 5 ml nuclei isolation buffer (0.1% triton X-100, 125 mM sorbitol, 20 mM potassium citrate, 30 mM $MgCl_2$, 5 mM EDTA, 5 mM 2-mercaptoethanol, 55 mM HEPES at pH 7.5) with 1X protease inhibitor, and transfer into a 7 mL Tenbroeck potter. Grind 10 times slowly on ice; then transfer the solution into in several 2 ml VK05 tube, homogenized by Precellys Evolution beads homogenizer (Bertin technologies, 7800 rpm, 30s each time, 20s pause each grinding cycle, repeat 5 times). Over 1 million

nuclei were isolated and digested overnight by Dpn II, DNA ends were labeled with biotin-11-dCTP (Jena Bioscience, cat. no. NU-809-BIOX-L) at 22°C for 4h in thermomixer (shake at 900 rpm, alternating 30 seconds on and 4 minutes off), then ligated by T4 DNA ligase (Thermo Scientific, cat. no. EL0012) at 22°C for 4h. The purified Hi-C DNA was sheared for 60 seconds by covaries E220 evolution (Peak incident power:175W, duty cycle: 10, intensity: 190, cycles per burst: 200) and libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit (NEB, cat. no. E7645) following the standard protocol, DNA concentration was quantified using a Qubit™ 4 Fluorometer (Invitrogen), and the average size of the library was detected by Bioanalyzer (Agilent Technologies), the final library was sequenced with 150 bp paired-end reads on an Illumina HiSeq 2000 platform at MPI and Novaseq X at Azenta. To check library quality and adjust sequence depth, an aliquot of library was sent for test sequencing. Two biological replicates were performed for each strain.

**High-Molecular-Weight (HMW) genomic DNA extraction and Nanopore sequencing**

HMW gDNA was extracted from *U. pinnatifida* strains, *D. dichotoma* female strains and *D. dudresnyi* strains using the NucleoBond® HMW DNA kit (Macherey-Nagel, cat. no. 740160.20) in combination with Lysis Buffer CF (Macherey-Nagel, cat. no. 740946). Approximately 300 mg of frozen algal tissue was ground in liquid nitrogen using a pre-chilled mortar and pestle. The tissue was grinded into fine powder and suspended in Lysis Buffer CF, gently inverted to ensure homogenous mixing, followed by incubation at 37 °C for 10 minutes. Subsequently, the following reagents were added sequentially, with gentle inversion after each addition: 5 mL of 5 M NaCl, 200 µL of 0.5 M EDTA, 4 mL of 10% CTAB in 0.7 M NaCl, 200 µL of 10% Triton X-100, 400 µL of Proteinase K (20 mg/mL), and 200 µL of RNase A. The extraction process then followed the manufacturer's protocol for HMW DNA isolation. DNA concentration was quantified using a Qubit™ 4 Fluorometer (Invitrogen), and fragment size distribution was assessed using the FEMTO Pulse system (Agilent Technologies) to confirm the integrity of HMW gDNA. For long reads sequencing, 1 µg of HMW gDNA was used to prepare libraries following the standard protocol of the Oxford Nanopore Technologies (ONT) ligation sequencing kit SQK- LSK110 (*D. dichotoma, D. dudresnayi*) and SQK-LSK114 (*U. pinnatifida*). Sequencing was performed on MinION R9.4 (*D. dichotoma* and *D. dudresnayi*) and PromethION2 R10.4 (*U. pinnatifida*) platforms. Basecalling was performed by ONT dorado v0.3.4 (https://github.com/nanoporetech/dorado).

***De novo* genome assembly and scaffolding**

ONT long reads draft assemblies were performed by canu[92], flye[93], NextDenovo[94] and Shasta[95] long read assemblers with default settings. The best draft assembly for each species was chosen based on the contig fragmentation level, N50, repeated sequences behavior, the re-mapped read coverage and the overall assembly length. We improved nanopore-based draft genome assemblies to chromosome-level resolution using Hi-C data. Specifically, *C. linearis*, and *S. ischiensis* draft assemblies were scaffolded using the 3D-DNA pipeline,[96] while the *D. dichotoma* female assembly was improved using the HapHiC pipeline (v1.0.7)[97]. Mapping in situ Hi-C data to existing *D. herbacea* male and female genomes, as well as to *D. dichotoma*, revealed multiple mis-assemblies

in the reference scaffolds. To address these, we applied the 3D-DNA pipeline to the draft assemblies of *D. herbacea* male, female and *D. dudresnayi*[26] and the HapHiC pipeline (v1.0.7) to the *D. dichotoma* male genome[26] using Hi-C datasets generated in this study. Mis-assemblies and chromosome rearrangements were corrected manually using Juicebox[45]. Overall, we obtained near telomere-to-telomere (T-to-T), chromosome-level assemblies for all species included in this study. In the updated assemblies, chromosomes were renamed and artificially oriented according to previous reference versions. Genome completeness was assessed using BUSCO (v5.8.2) with the eukaryota_odb10 dataset[38].

**Genome annotation and macrosynteny analysis**

Gene annotations were lifted over from the previous gene models to the new assembled genomes used Liftoff (v1.6.1)[98], and repeat was re-annotated by Earl Grey[99] and manually curated with *Ectocarpus* sp7 TE library[40]. Homologous syntenic blocks were built with GENESPACE[41] R package and visualized by LINKVIEW2[100].

**Hi-C data processing**

Hi-C data was processed by Juicer[101] pipeline with default parameters, Hi-C reads were mapped to chromosome-level assemblies of each species with BWA, after alignment, chimera handing, merge, sort and removed duplicate reads pairs, juicer format contact maps(.hic) were generated with resolutions of 5kb, 10kb, 25kb, 50kb, 100kb and 250kb and visually inspected in Juicebox[45]. HiCRep[42] was used to assess the Hi-C data reproducibility between replicates with stratum-adjusted correlation coefficient (SCC) method at resolutions of 5, 10, 25, 50, 100 and 250kb. The SCC scores were averaged across chromosomes. Biological replicates were merged to get a higher resolution Hi-C matrix. Juicer format pair files 'merged_nodups.txt' was convert to pairs with 'merged_nodup2pairs.pl'. The distance law represents decay of the average contact frequency was calculated directly from pair files with HiContacts[43] R package, the most informative genomic distance from 10 kb to 1 Mb was plot for each sample separately.

**Compartment A/B**

Chromatin compartments were inferred from based on the plaid interaction patterns in Hi-C contact matrices, which reflect spatial segregation of A and B compartment (open and close chromatin). To annotate compartment A/B, juicer format matrix'.hic' files with a mapping quality score (MAPQ) ≥ 30 were converted to .cool format by hic2cool (https://github.com/4dn-dcic/hic2cool, v1.0.1) and matrix balancing was performed using the cooler (v1.0.1) [102] by iterative correction and eigenvector (ICE) method. Eigenvector decomposition was conducted using the cooltools[103] (v0.4.0) eigs-trans function to calculate the first principal component (E1) at multiple resolutions. The sign of the E1 values was corrected based on their correlation with genomic features: regions with higher GC or gene density were designated as compartment A (positive E1), while regions with lower density were assigned to compartment B (negative E1). To refine compartment annotations, E1 signal tracks were visually inspected alongside Hi-C contact maps with KR normalization[104] in Juicebox[45]. Manual validation focused on ensuring concordance between compartment

signals and local intra- and inter-chromosomal interaction patterns. Saddle plots were generated following the instruction (https://cooltools.readthedocs.io/en/latest/notebooks/compartments_and_saddles.html), 95% of the genome was divided into 38 groups based on digitized eigenvector values, and we calculated the average observed/expected contact frequencies in each pair of group[103]. From each saddle plot, we extracted the average interaction frequencies within B–B compartments (top left) and A–A compartments (bottom right). Compartment strength was quantified as the ratio of (AA + BB) / (AB + BA)[103].

**Insulation score and boundary calculation**

Insulation score reflects the aggregation level of contact frequency map, high insulation is classified as chromatin boundaries in plant and animal[105]. To annotate insulation, cooltools[103] (v0.4.0) was used with balanced cooler format matrix'.mcool' files as input. Using Hi-C contacts matrix with a mapping quality score (MAPQ) $\geqslant$ 30 at 10 kb resolution, we tested multiple window sizes (50, 100, 200, 300, 400, and 500 kb). Based on visual inspection with Hi-C matrix in Juicebox[45], the 100 kb window size was selected for downstream analysis. Strong and weak boundaries were identified using the parameters '--threshold Li --min-dist-bad-bin 2'.

**Identification of centromeric retrotransposons**

Each genome was first queried via tblastn searches[106] using the ECR-1 pol protein from the *Ectocarpus* sp. 7 centrophilic retrotransposon[31]. The hits in each genome were then manually curated and consensus sequences were produced for each corresponding family following the method of Goubert et al. [107]. Briefly, the nucleotide sequences were extracted with flanking sequence, clustered based on similarity, and then reduced to a single full-length consensus sequence based on alignments that extended from the start of the left long terminal repeat to the end of the right long terminal repeat.

As a second approach, broad centromeric regions of ~1 Mb were visually defined from the Hi-C contact maps for each species (see Fig. 4). As performed previously for *Ectocarpus* sp. 7 [31], putative centromeric repeats from the automated Earl Grey repeat libraries were defined based on i) their enrichment in the centromeric regions relative to non-centromeric regions, and ii) their presence in multiple, if not all, centromeric regions. Any other repeat families beyond the *ECR* families that met these criteria were also manually curated following the same process, which included the centrophilic Copia families from *D. dictyota* and *S. ischiensis*.

If present, redundant repeat models from the automated Earl Grey libraries were replaced with the manually curated models, and the updated libraries were then used to annotate repeats against the relevant genome using RepeatMasker v4.1.6 (repeatmasker.org). The *Ectocarpus* sp. 7 genome was instead masked using the available curated library[40]. Additional satellite repeats were identified in each genome using Tandem Repeats Finder v4.09.1 and the parameters "2 7 7 80 10 50 2000", followed by extracting repeats with monomer lengths > 10 bp. If TE and satellite repeat annotations overlapped, TEs were given precedence. The genomic densities of TEs (including

centrophilic families and related *ECR-like* families) and satellites were then calculated in nonoverlapping sliding windows.

For *ECR* and *ECR-like* families, gag and pol protein sequences were manually extracted from the nucleotide consensus sequences and aligned using mafft v7.525 and the "L-INS-i" model[108]. Alignment gaps were filtered out using trimAl v1.4rev22 and the model "gappyout"[109]. A maximum likelihood phylogeny was then produced using IQ-TREE v2.3.0 with ultrafast bootstrapping and ModelFinder ("-bb 1000 -m MFP")[110]. For Copia, Gag and Pol protein sequences were extracted from the consensus sequences for the centrophilic families and for all curated *Ectocarpus* sp. 7 families[40]. The *D. dichotoma* reverse transcriptase domain was used to perform tblastn searches against the automated Earl Grey repeat models for each species, corresponding protein sequences were extracted and a preliminary phylogeny was produced using the method above. This phylogeny was used to identify a candidate set of Copia families that are closely related to the centrophilic families, and full-length gag and pol sequences were extracted for each of these families (with manual curation performed if required to extend consensus sequences or correct for frame-shifts). A final phylogeny using the resulting gag and pol protein sequences was then produced using the same method as for *ECR*/*ECR-like* elements, with more distantly related Copia families from the curated *Ectocarpus* sp. 7 library forming the outgroup.

### Ancestral genome reconstruction

Ancestral genome reconstruction was performed using AGORA[60] and the following genomes and annotations: *Ectocarpus* sp. 7[31], *C. linearis*, *U. pinnatifida*, *D. herbacea*, *D. dictyota* (all this study), and *S. promiscuus* and *F. serratus* [25]. Single copy orthologs present in all genomes (n=4,847) were defined using OrthoFinder v2.5.5 ("-S diamond_ultra_sens")[111]. Contiguous Ancestral Regions (CARs) flanking the centromeres were then extracted and the genes present in these CARs were mapped onto the chromosomes (see Figure 5).

Rearrangements were manually curated using the GENESPACE macrosynteny analysis[41]. A second OrthoFinder run including *S. ischiensis* proteins was performed, and orthologs between *D. herbacea* and *S. ischiensis* were extracted. One-to-many and many-to-many orthology relationships were retained if all of the genes were on the same chromosome in both species. The *S. ischiensis* genes were then colored according to the chromosome of their ortholog(s) in *D. herbacea*.

## Acknowledgements

We thank Remy Luthringer, Masakazu Hoshino, Kenny A Bogaert, Andrea Belkacemi, Dorothe Koch and Anagha Kerur, for assistance with algal cultures; Jaruwatana Sodai Lotharukpong and Romy Petroll for help with phylogenetic trees; Jeromine Vigneau, Michael Borg, Erica Dinatale and Zhigui Bao for discussions.

## Funding

This research was funded by the Max Planck Society, European Research Council, grant 864038 (SMC), the Gordon and Betty Moore Foundation (SMC). RC was recipient of a Marie Skłodowska-Curie Postdoctoral Fellowship (grant 101109906). Pengfei Liu is thankful to the International Max Planck Research School 'From Molecules to Organisms.'

## Author contributions

Conceptualization: SMC. Methodology: SMC, RC, PL, CL. Investigation: PL, RL, FBH, RC. Visualization: PL, RC, FBH. Funding acquisition and project administration: SMC. Supervision: SMC, CL, FBH. Writing (original draft): SMC, PL, RC. Writing (review & editing): SMC.

**Competing interests:** Authors declare that they have no competing interests.

**Data and materials availability:** All data are available in the main text or the supplementary materials.

# Supplemental Figures

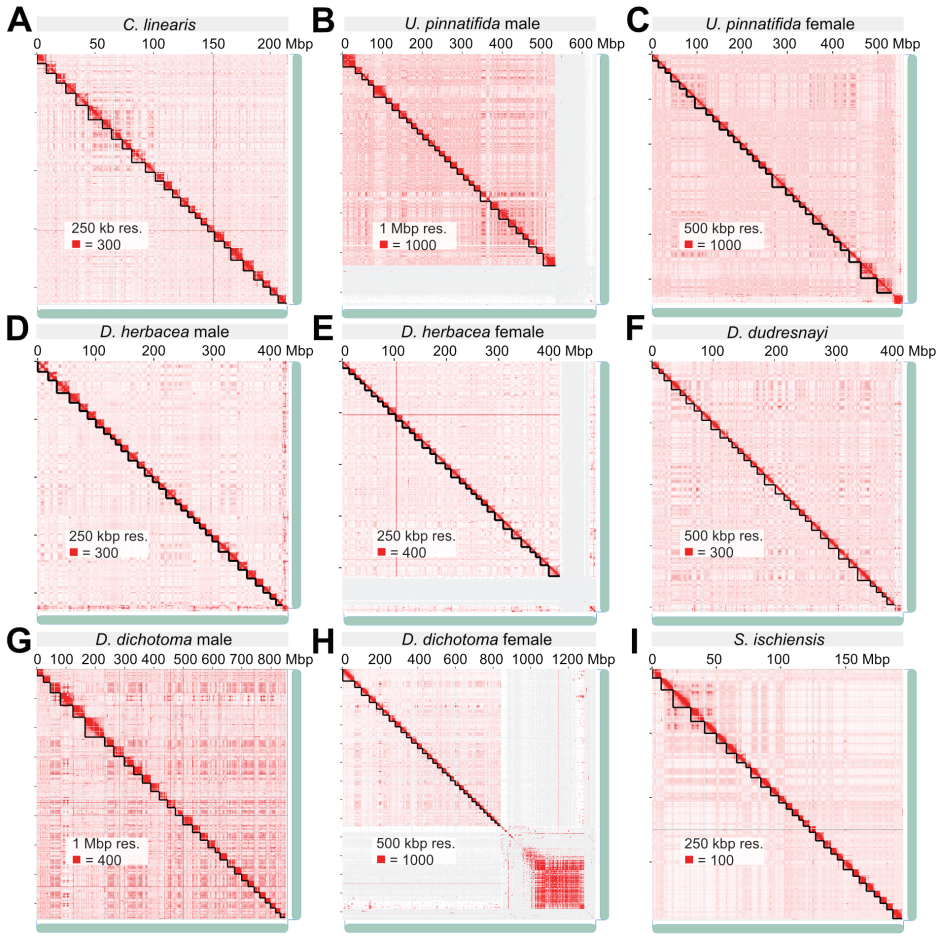



**Fig. S1.** (**A-I**) Genome assemblies for each strains by nanopore long reads and Hi-C reads, chromosomal boundaries and prominent interactions of telomere clustering and centromere clustering are clear seen in Hi-C maps for each species in Juicebox[45].

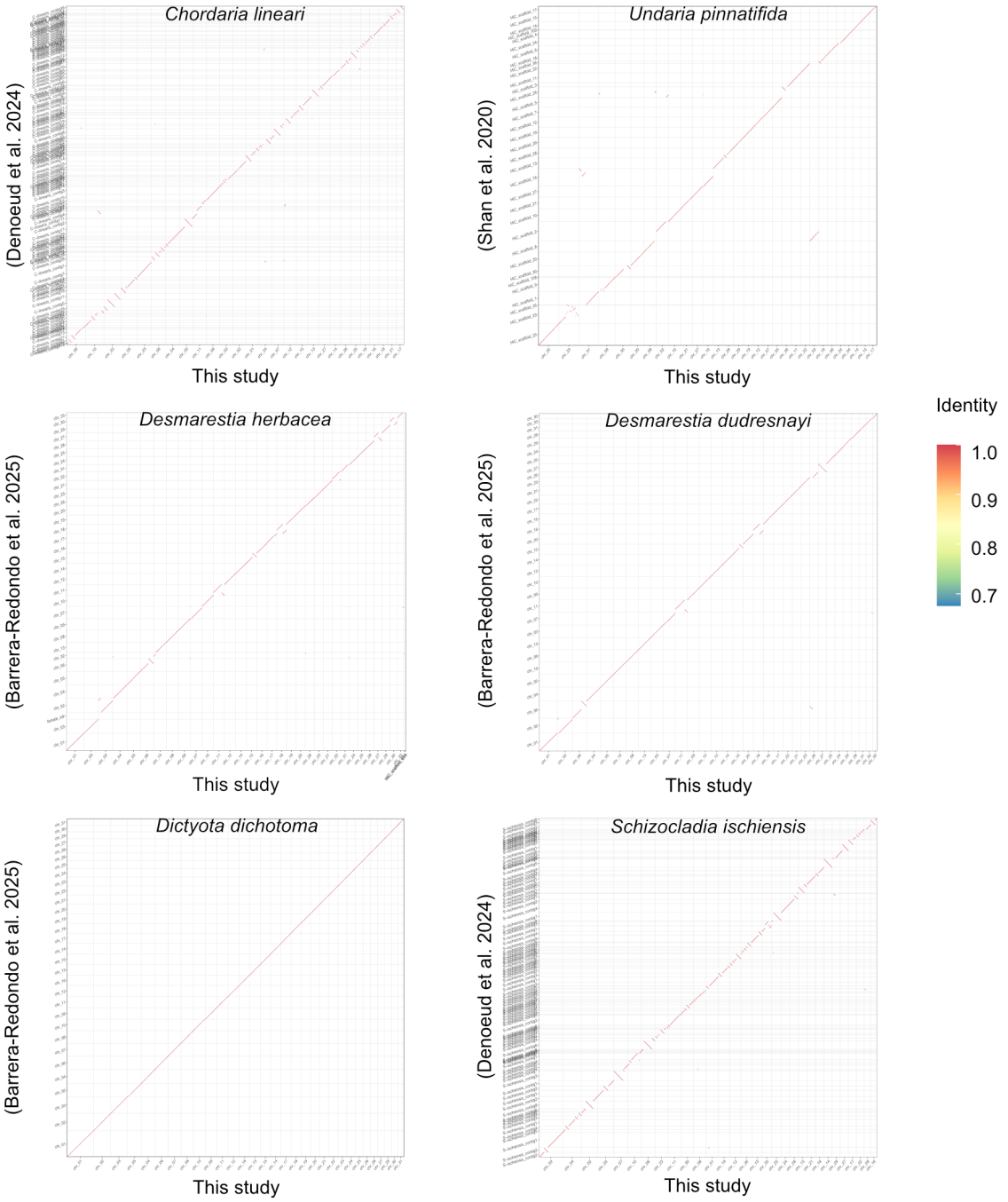



**Fig. S2.** Dotplot comparison of previous and current genome assemblies. Whole-genome pairwise alignments were performed using minimap2[112], and the resulting alignments were visualized with paf2dotplot (https://github.com/moold/paf2dotplot). The generated dotplots indicate high sequence identity between the old and new assemblies

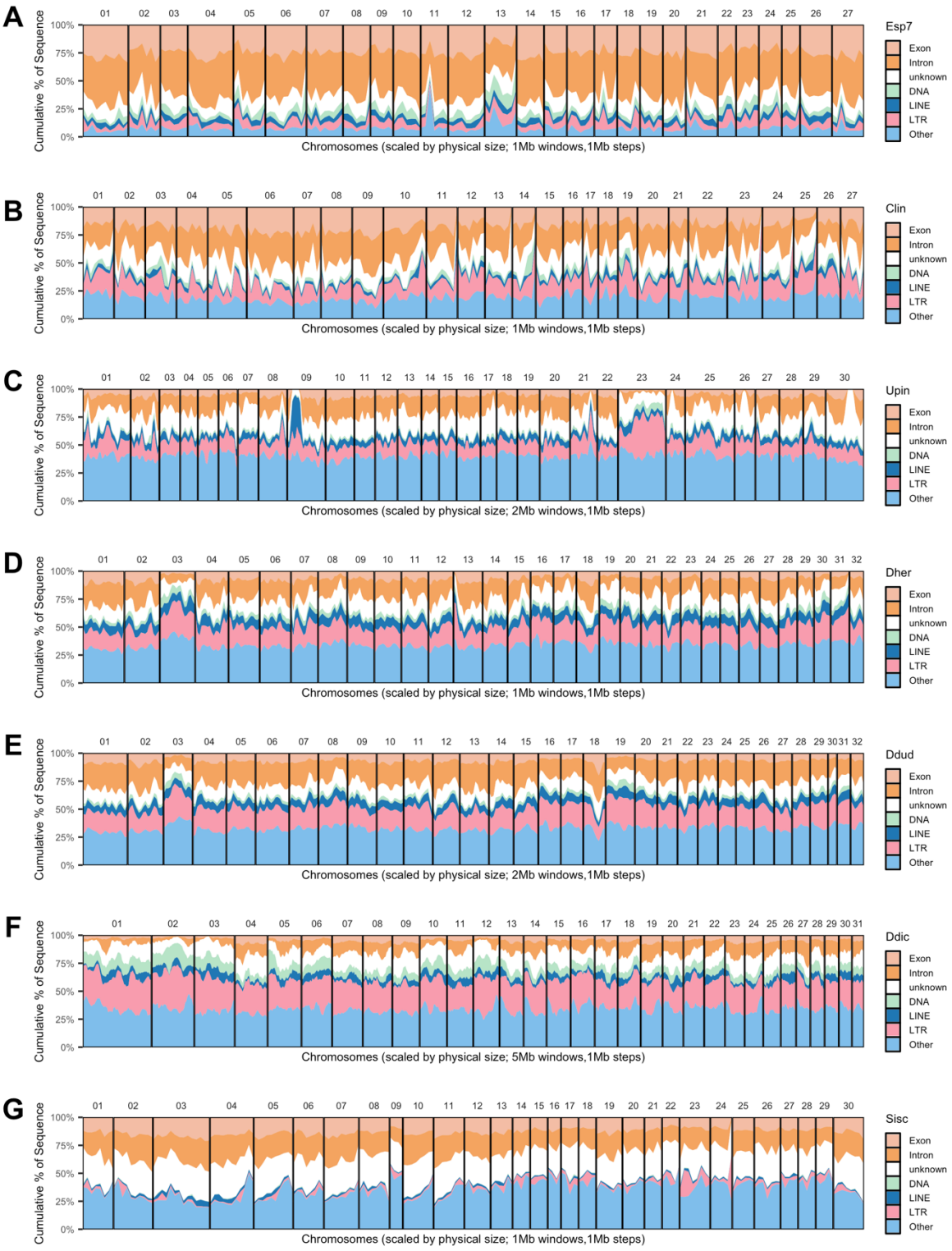



**Fig. S3.** (**A-G**) Landscape of exon, intron, and TE families (DNA, LINE, LTR and others, others include simple repeats, satellite, unclassified DNA, etc) for each species. (Esp7 = *Ectocarpus* sp. 7, Clin = *Chordaria linearis*, Upin = *Undaria pinnatifida*, Dher = *Desmarestia herbacea*, Ddud = *Desmarestia dudresnayi*, Ddic = *Dictyota dichotoma*, Sisc = *Schizocladia ischiensis*).

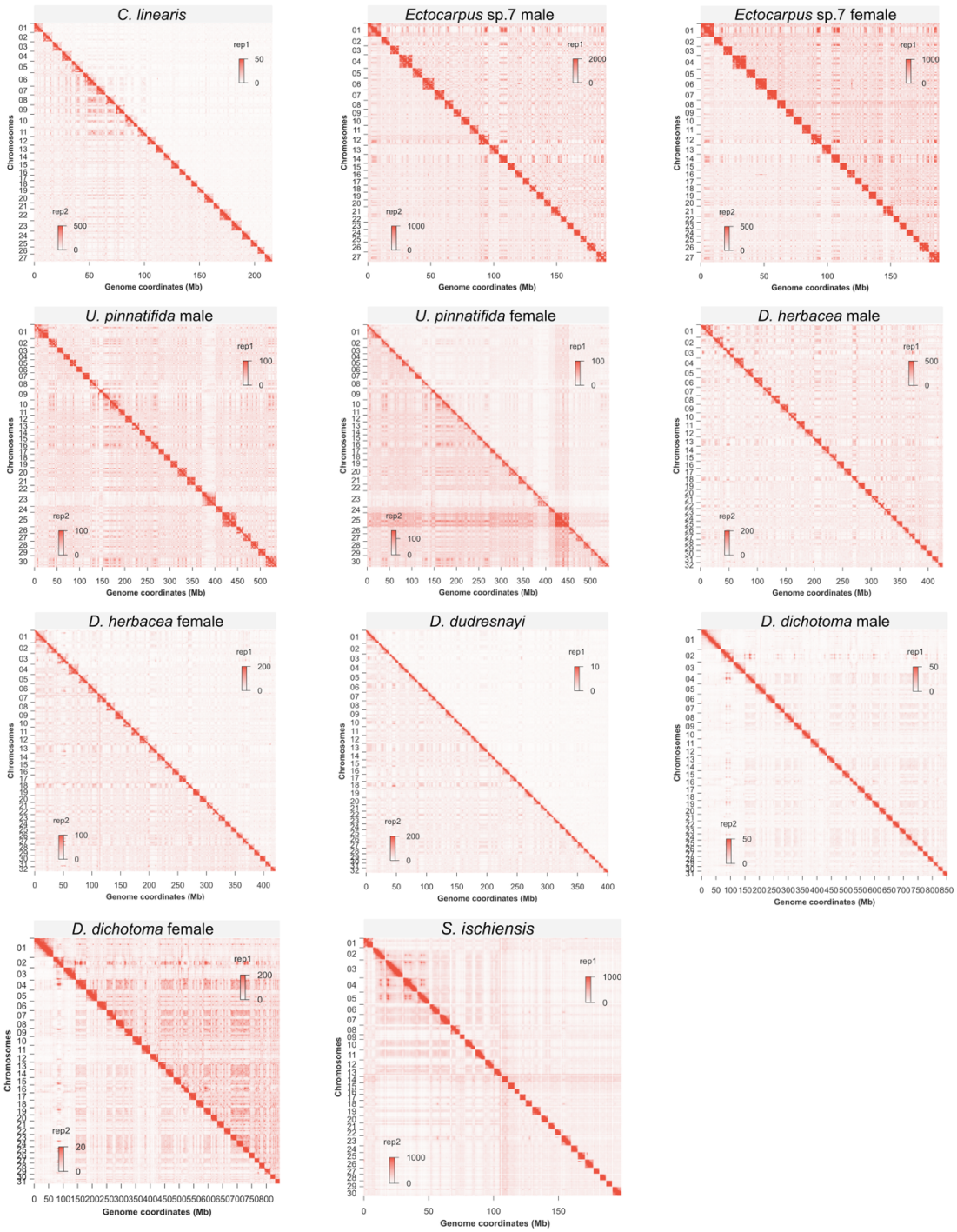



**Fig. S4**. Biological replicates of whole genome-wide Hi-C maps for each species visualized in Juicebox[45] at 250k resolution

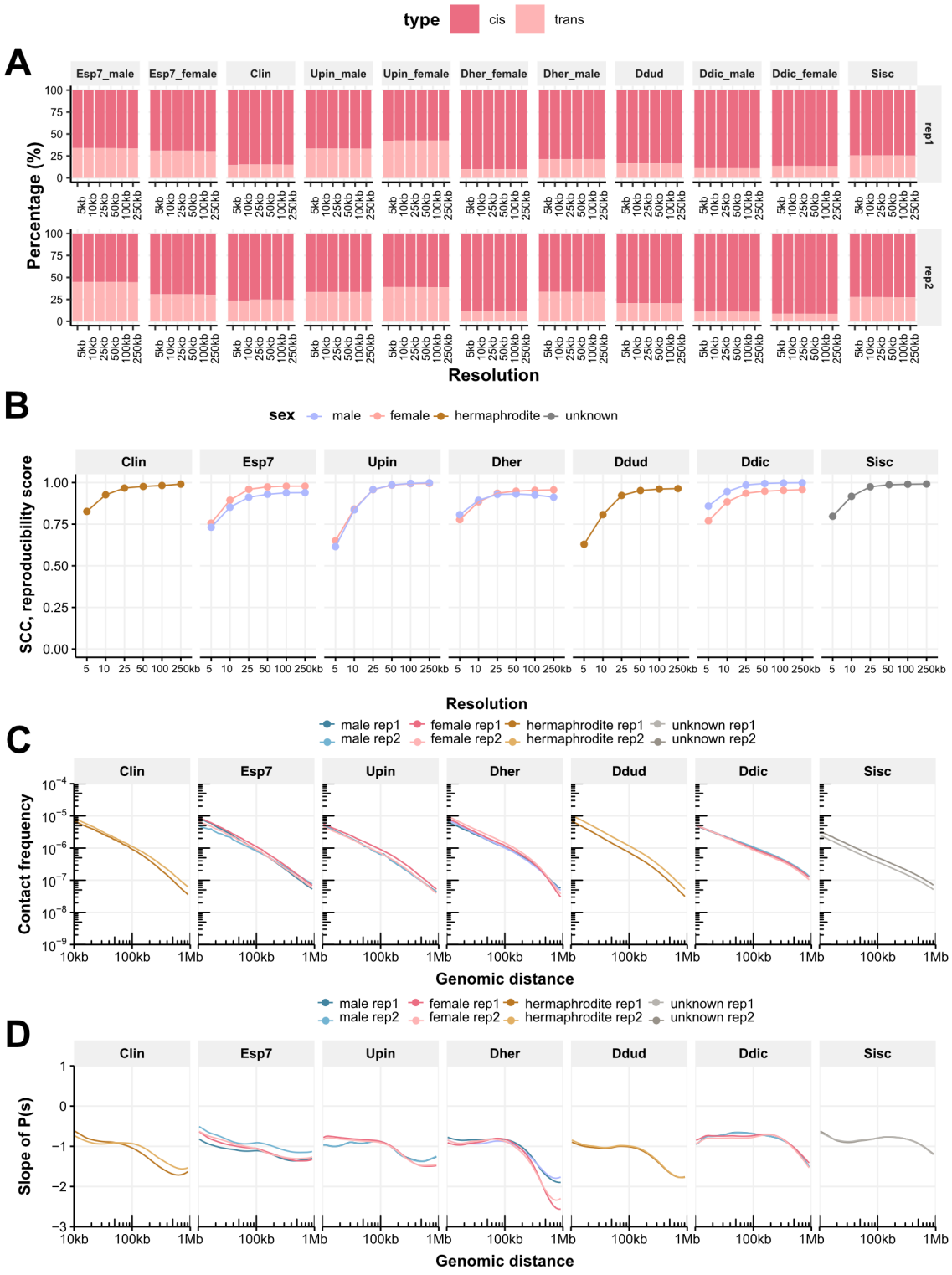



**Fig. S5.** Hi-C data statistics. (**A**) Percentage of cis and trans interactions at different resolutions. (**B**) corrections score between biological replicates, and (**C**) Interaction frequency curve of distance law and (**D**) slope of distance law for each replicate of each species

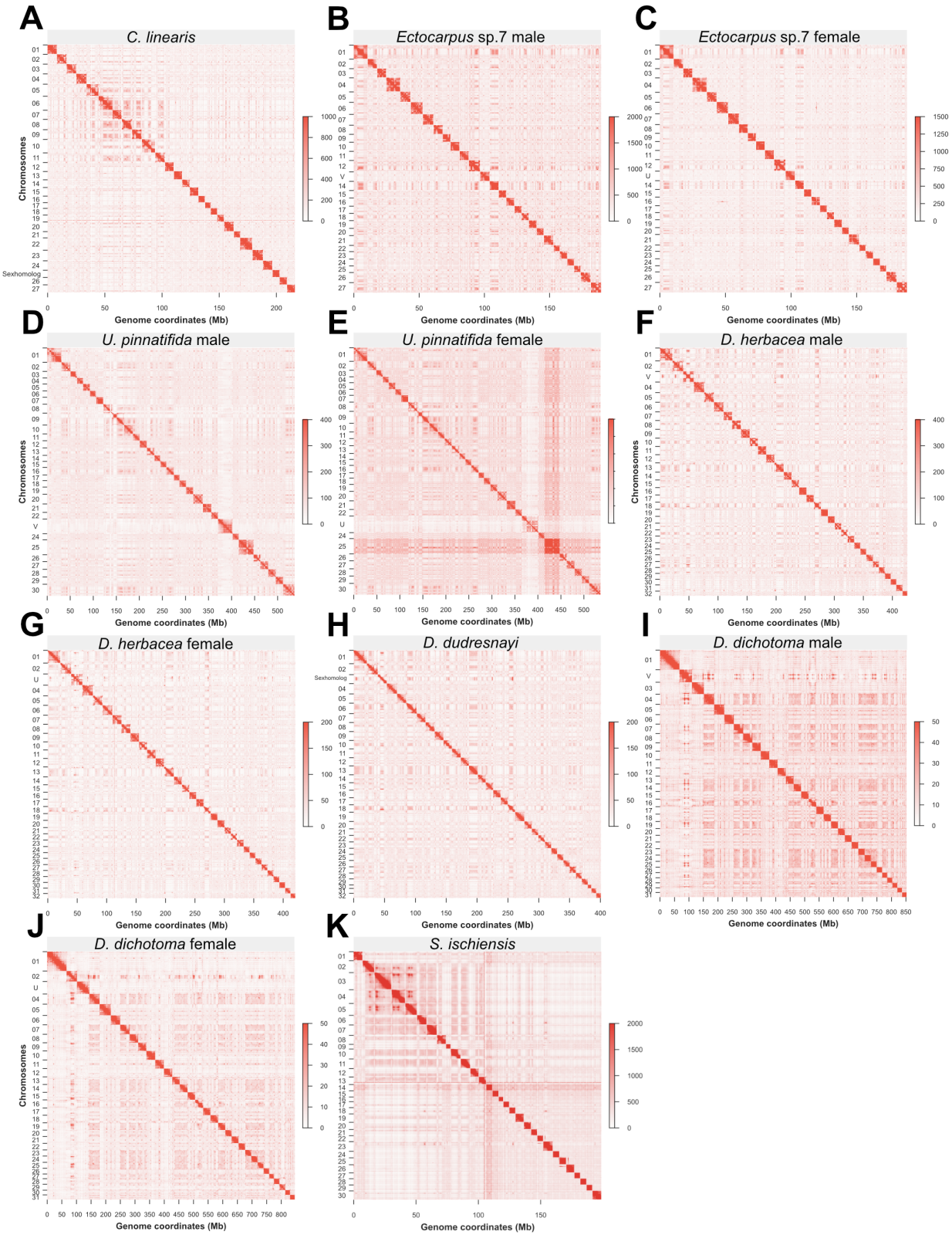



**Fig. S6.** Final whole genome-wide Hi-C maps for each species at 250k resolution with merged Hi-C data.

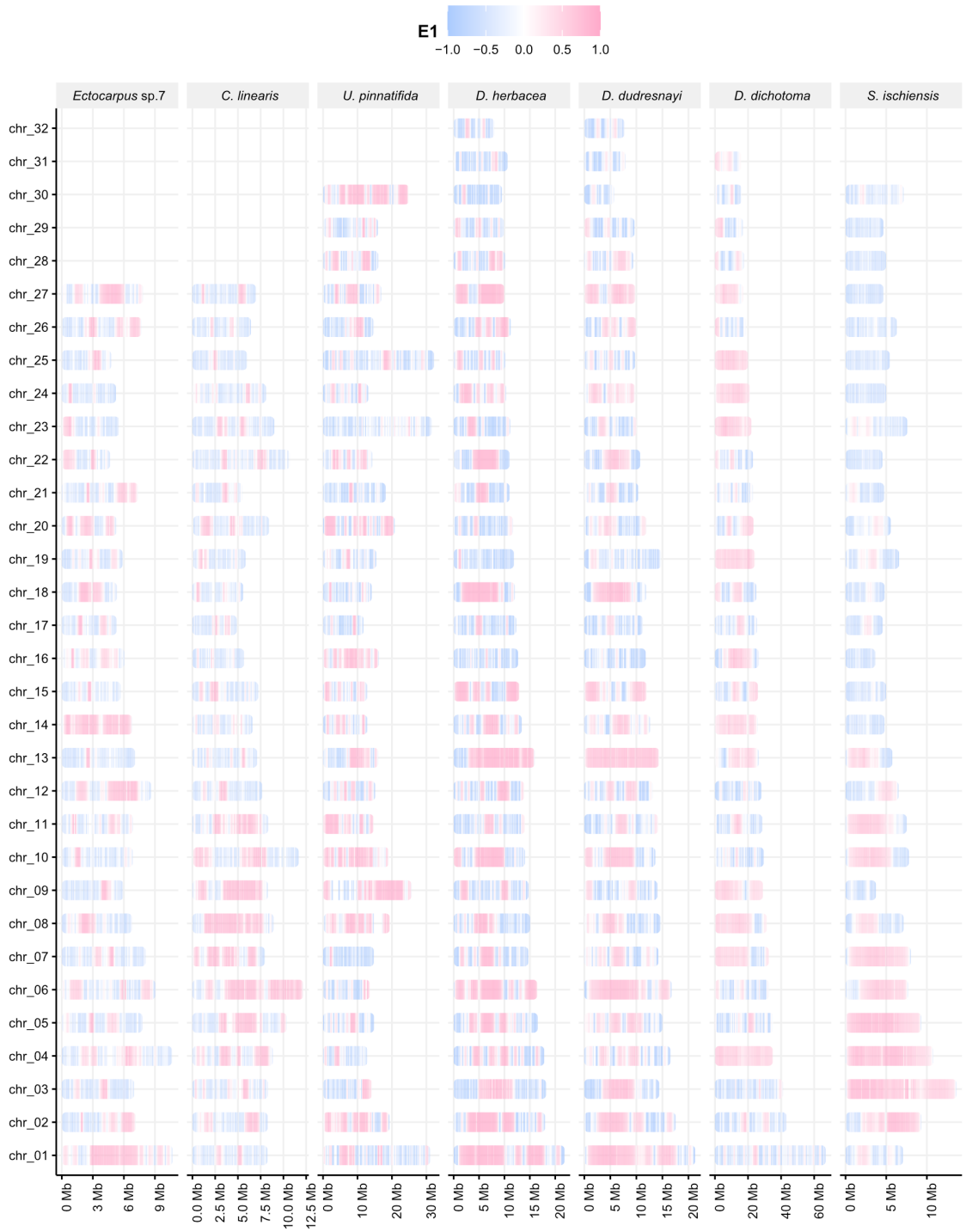



**Fig. S7**. A/B compartment distribution on each chromosome for each species (E1 is the first eigenvector value).

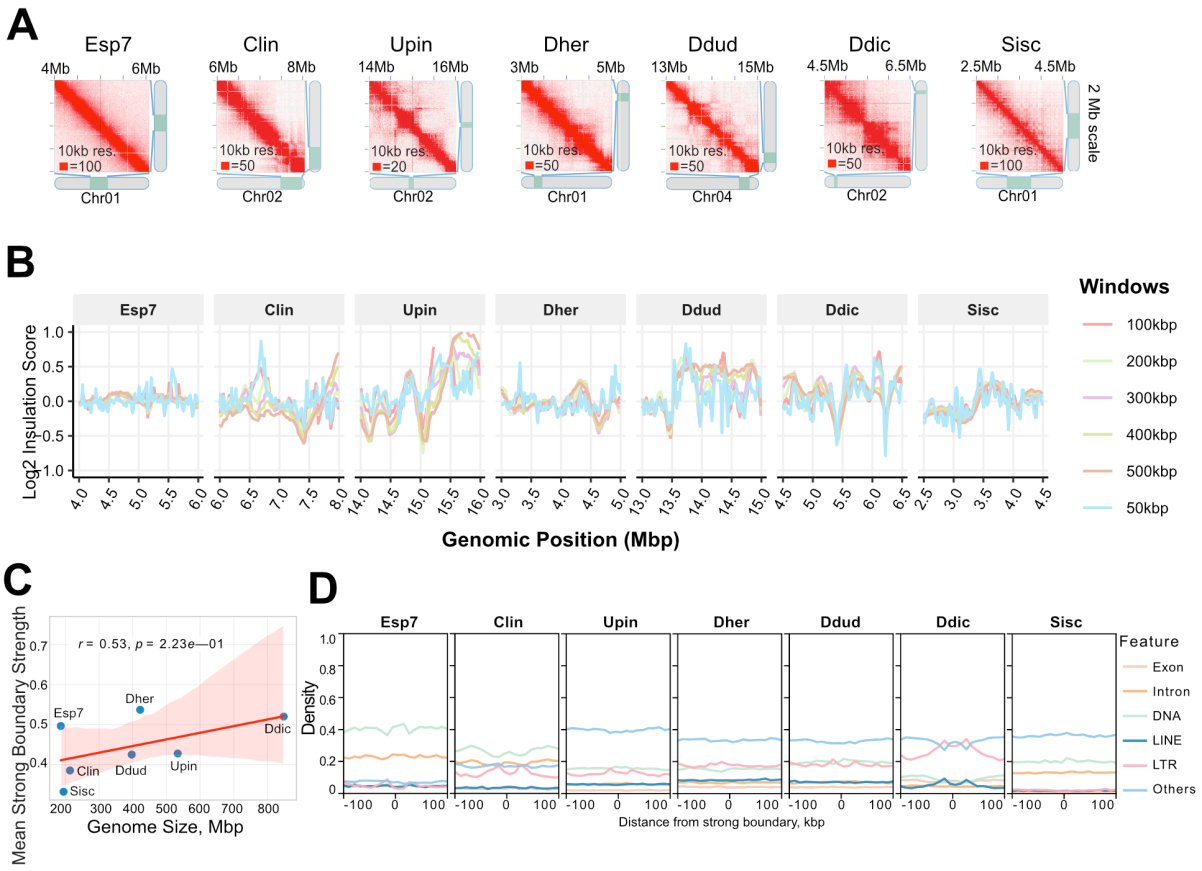



**Fig. S8**. Insulation and boundaries profiles across species. (**A**) A 2 Mb-scale Hi-C matrix for each species. (**B**) Insulation scores were calculated using 10 kbp resolution Hi-C contact matrices with multiple sliding window sizes (50 kbp, 100 kbp, 200 kbp, 300 kbp, 400 kbp, and 500 kbp). To compute insulation profiles, a diamond-shaped window was slid along the genome diagonal, with one corner anchored on the main diagonal of the contact matrix. For each genomic position, the sum of contacts within the window was calculated, with lower scores indicating stronger insulation boundaries. (**C**) Correlation between extracted mean boundary strength score and genome size, the shaded region represents the 95% confidence interval around the regression line. (**D**) Exon, intron and TEs enrichment on boundaries upstream and downstream at 10k window size. Numbers below brackets represent the number of boundaries. The lower and upper hinges of the box correspond to the first and third quartiles (the 25th and 75th percentiles). The upper whisker extends from the hinge to the largest and smallest values no further than 1.5x IQR from the hinge (Inter-Quartile Range, distance between the first and third quartiles). Significant was determinant by two-sample Wilcoxon rank sum test.

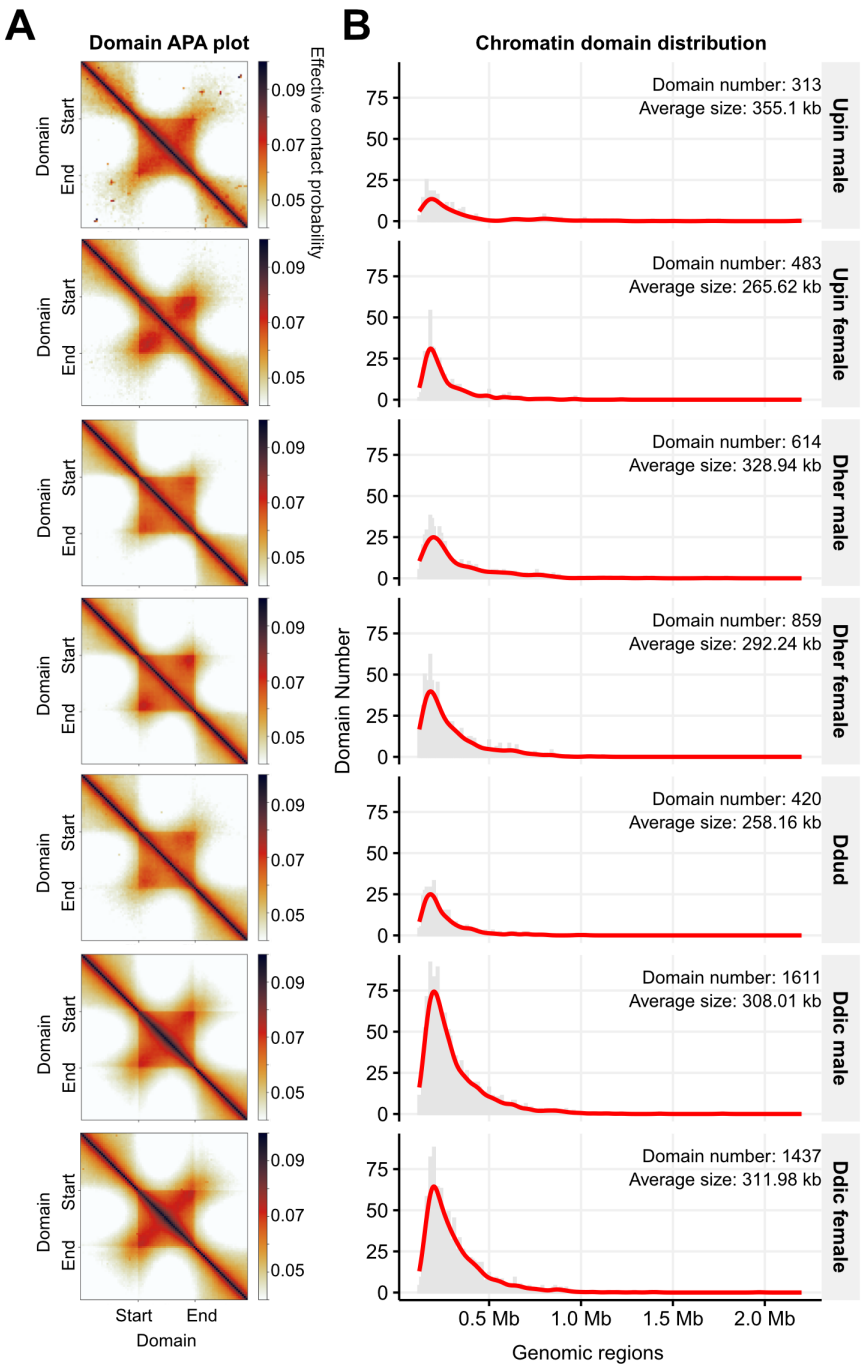



**Fig. S9**. Chromatin domains profile for species. (**A**) Aggregate chromatin domain plot of normalized Hi-C matrix at 10kb resolution by FAN-C[113]. For each TAD, the contact map from (start - length) to (end + length) was rescaled to a 90×90 matrix and averaged. The TAD body corresponds to the central pixels. (**B**) Domain size distribution in each species.

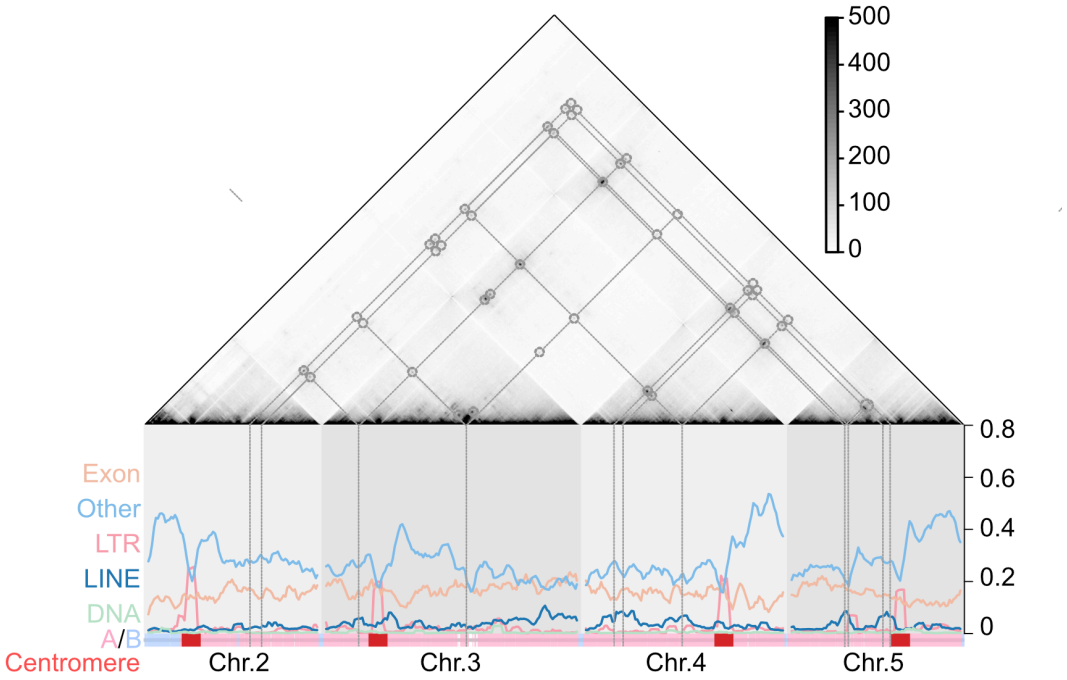



**Fig. S10**. Example shows prominent inter-chromosomal interactions of a sub Hi-C matrix of *S. ischiensis* at 25k resolution, profiled with fraction of TEs family (LINE, DNA, LTR and others), exon in 500kb window size, ideogram shows compartment A/B and centromere annotations.

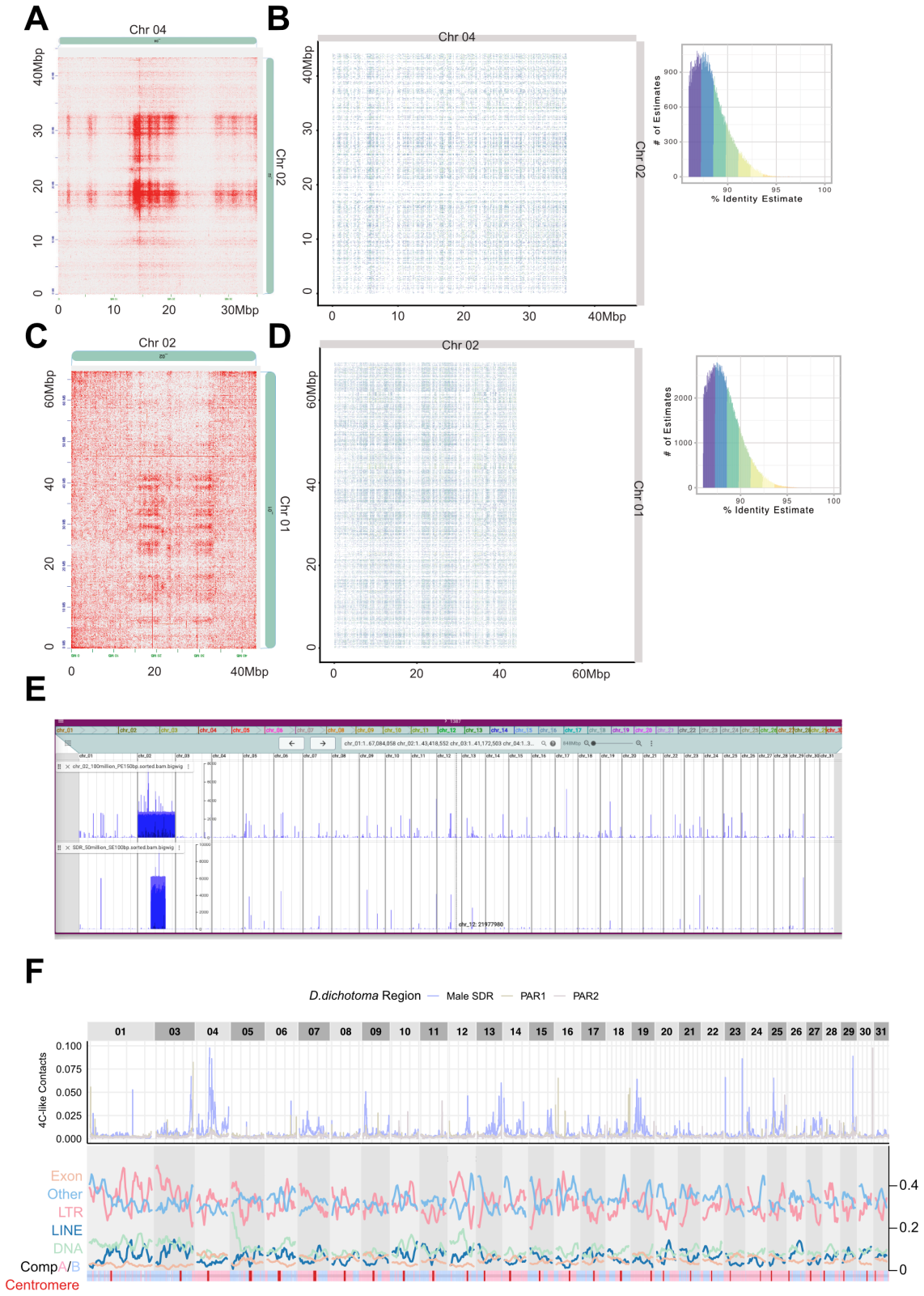



**Fig. S11**. Validation of long-range interactions between sex chromosomes (SDRs) and autosomes are not mapping artifacts and interaction patterns are not affected by TEs. (**A**) and (**C**) Examples showing strong inter-chromosomal

interactions involving the SDR on *D. dichotoma* male genome. (**B**) and (**D**) corresponding self-identify plots reveal low sequence similarity between the sex chromosomes and autosomes by ModDotPlot[114]. (**E**) Coverage of uniquely mapped reads: simulated 100 million paired-end (PE) 150 bp reads of the entire sex chromosome (upper), and simulated 50 million single-end (SE) 100 bp reads of male SDR (lower), shown on *D. dichotoma* male genome. (**F**) Genome-wide virtual 4C profile of contacts between PAR1, SDR, and PAR2 of sex and autosomes at 100kb Hi-C resolution and corresponded each TEs family density, compartment A/B and centromere (in red) profiles for *D. dichotoma* at 5Mb window size.

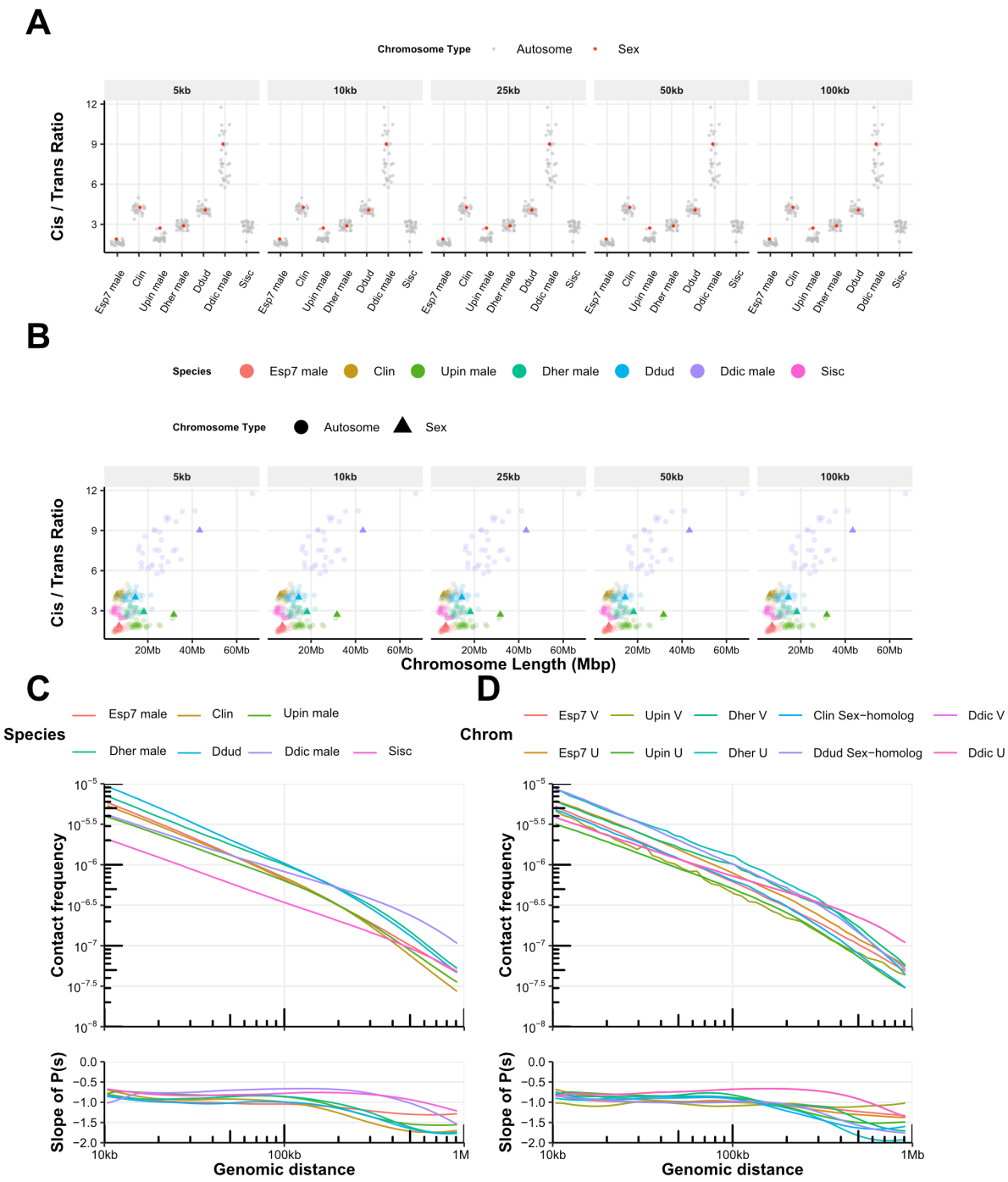



**Fig. S12**. Trans/cis interaction frequency ratio for (A) each chromosome and (**B**) as a function of chromosome size, shown for each species at resolutions of 5 kb, 10 kb, 25 kb, 50 kb, 100 kb, and 250 kb. (**C**) Distance-dependent interaction frequency (Ps) and corresponding slope for each species) and (**D**) sex chromosomes (Chr V and Chr U) and sex-homologs.

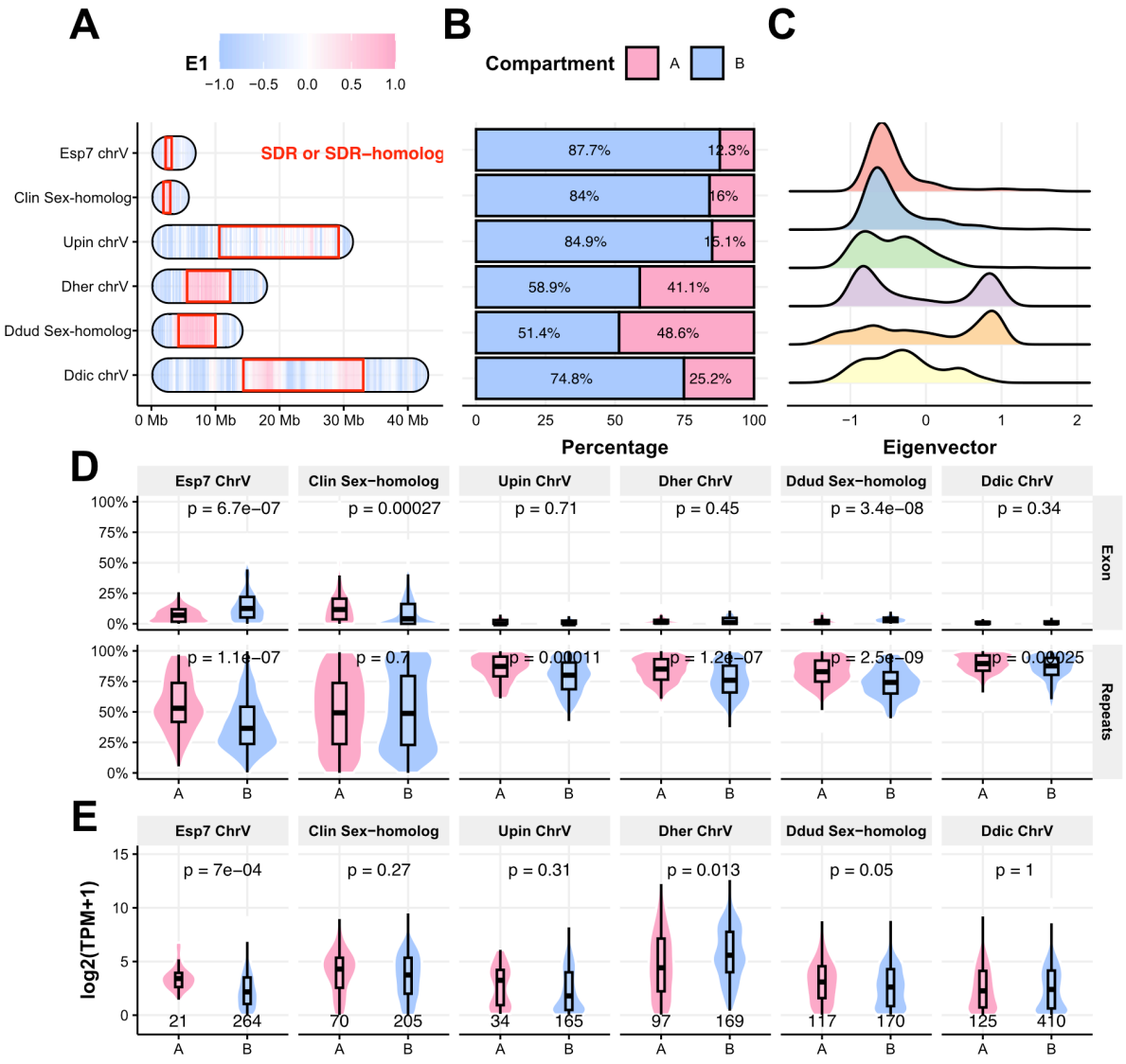



**Fig. S13**. Compartmentalization profile of sex chromosome (Chr V and Chr U) and sex homologs. (**A**) ideogram plot for eigenvector profiles, (**B**) percentage of A or B compartment, (**C**) eigenvector values distribution (**D**) exon and repeat density and (**E**) gene expression in A / B compartment for sex chromosomes and sex homologs of each species. Numbers below brackets represent the number of genes. The lower and upper hinges of the box correspond to the first and third quartiles (the 25th and 75th percentiles). The upper whisker extends from the hinge to the largest and smallest values no further than 1.5x IQR from the hinge (Inter-Quartile Range, distance between the first and third quartiles). Significance was determined by a two-sample Wilcoxon rank sum test.

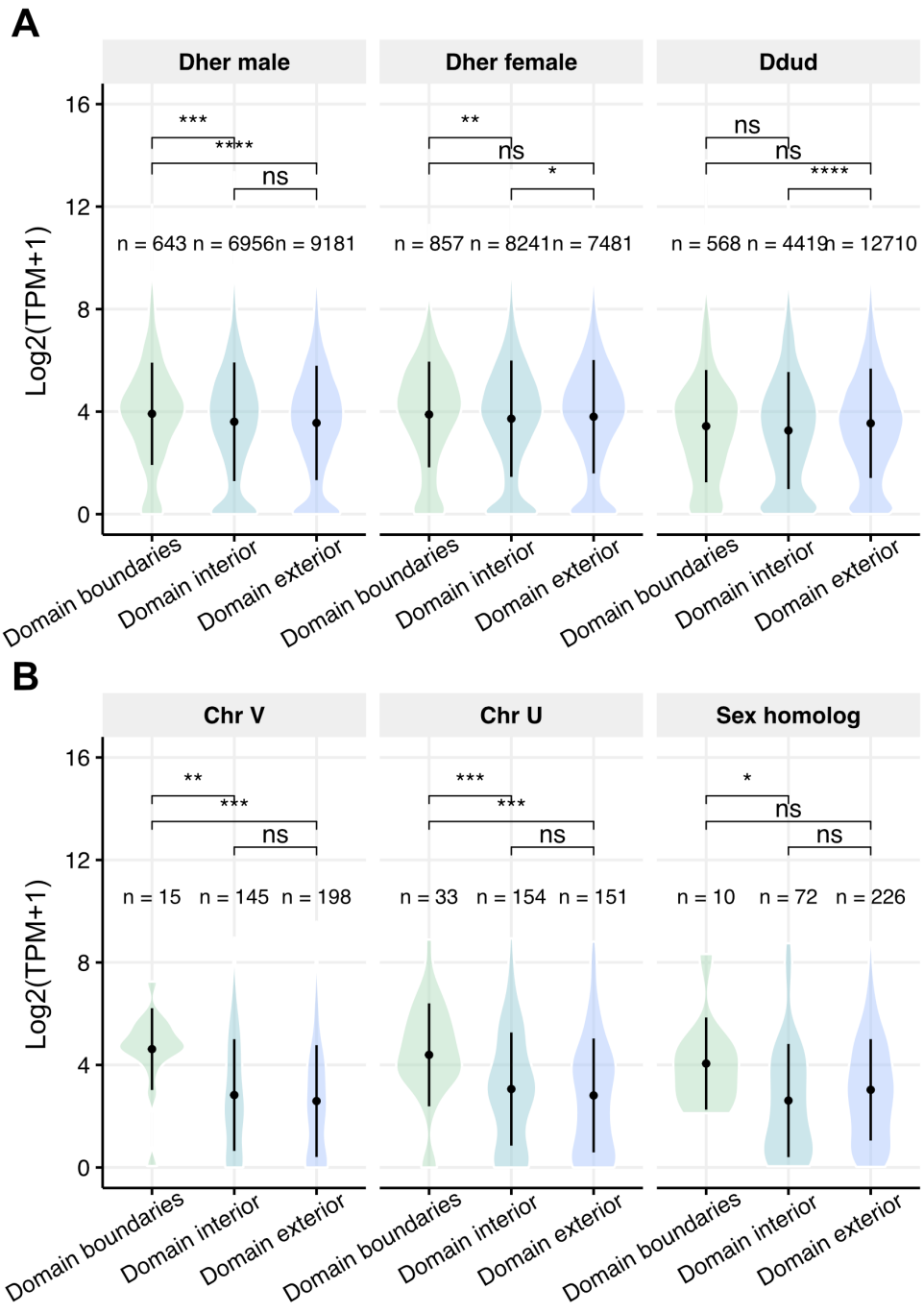



**Fig. S14**. Gene expression profile of chromatin domain for (**A**) whole genome *D. herbacea* (Dher) male female and *D. dudresnayi* (Ddud) and (**B**) sex chromosomes (Chr V and Chr U) and sex-homolog. Numbers above brackets represent the number of genes. Plots show mean ± standard deviation, with violins depicting full data distribution. Significance was determined by a two-sample Wilcoxon rank sum test (****: $p <= 0.0001$, ***: $p <= 0.001$, **: $p <= 0.01$, *: $p <= 0.05$, ns: not significant, $p > 0.05$).

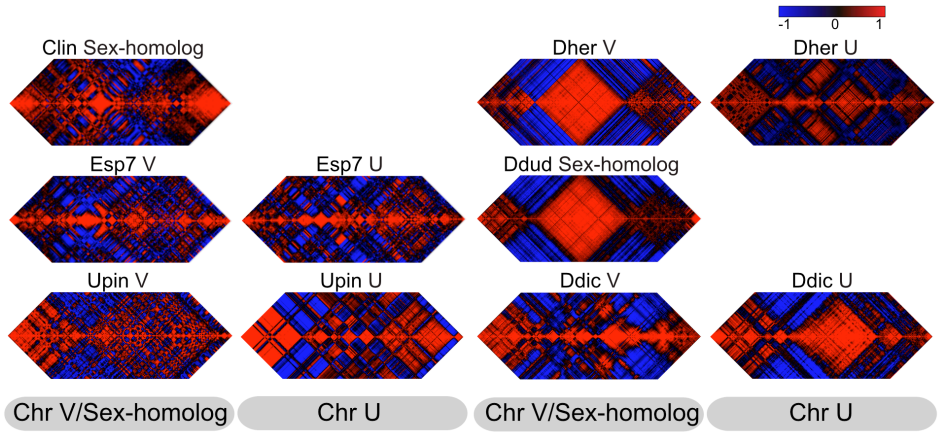



**Fig. S15**. Pearson correlation Hi-C maps at 50k resolution of sex chromosomes and sex-homologs for each species.

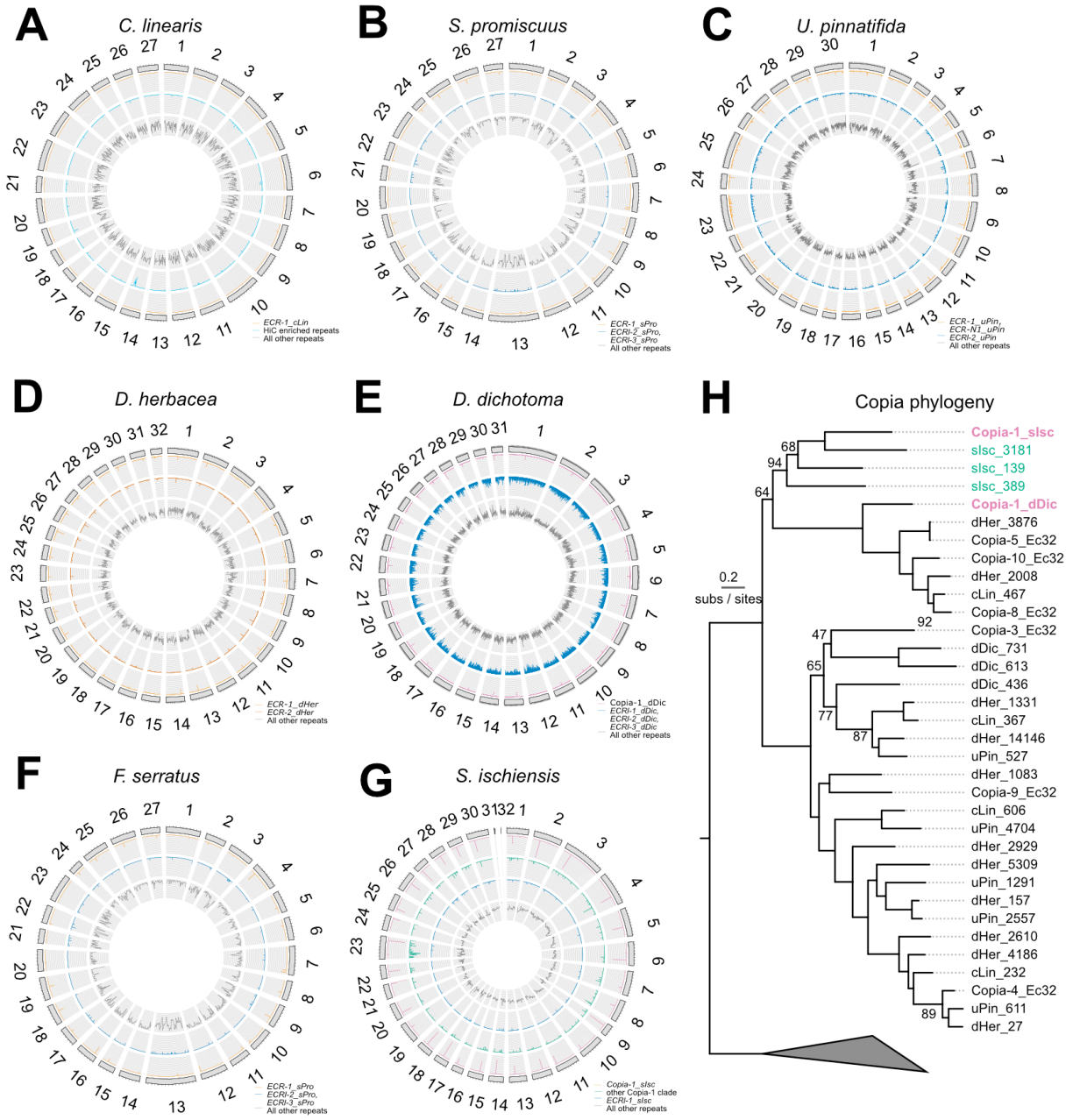



**Fig. S16**. Genome wide distributions of *ECR* and *ECR-like* LTR families and centrophilic Copia families (**A-G**). Genomic proportions for each specific repeat family are in 50 kb windows, whereas all other repeats are in 200 kb windows. Centrophilic retrotransposons have a characteristic single peak per chromosome. (**H**) Maximum likelihood phylogeny of gag and pol protein sequences for Copia families present in the genomes of the species analysed in this study (Ec32 = *Ectocarpus* sp. 7, cLin = *Chordaria linearis*, uPin = *Undaria pinnatifida*, dHer = *Desmarestia herbacea*, dDic = *Dictyota dichotoma*, sIch = *Schizocladia ischiensis*) build under the Q.pfam+R5 model. Centrophilic elements are colored purple and non-centrophilic elements in *S. ischiensis* are colored green (**G**). Ultrafast bootstrap values <95 are shown and the phylogeny is rooted on more distantly related Copia elements from *Ectocarpus* sp. 7.

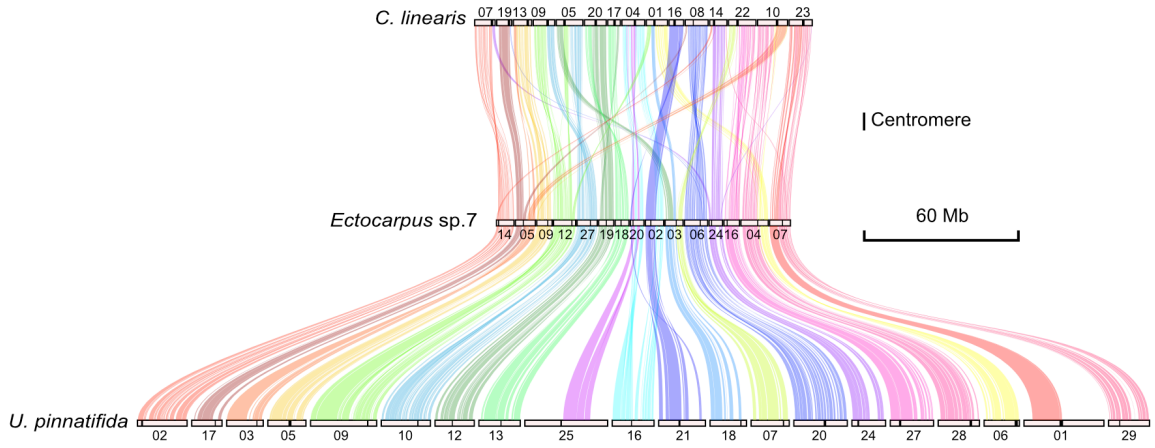



**Fig. S17**. Example shows inter-chromosomal rearrangements events found in *C. linearis*, *Ectocarpus* sp.7 and *U. pinnatifida*, synteny colors were rephased by genome of *U. pinnatifida*, centromeres are plotted as black bars.

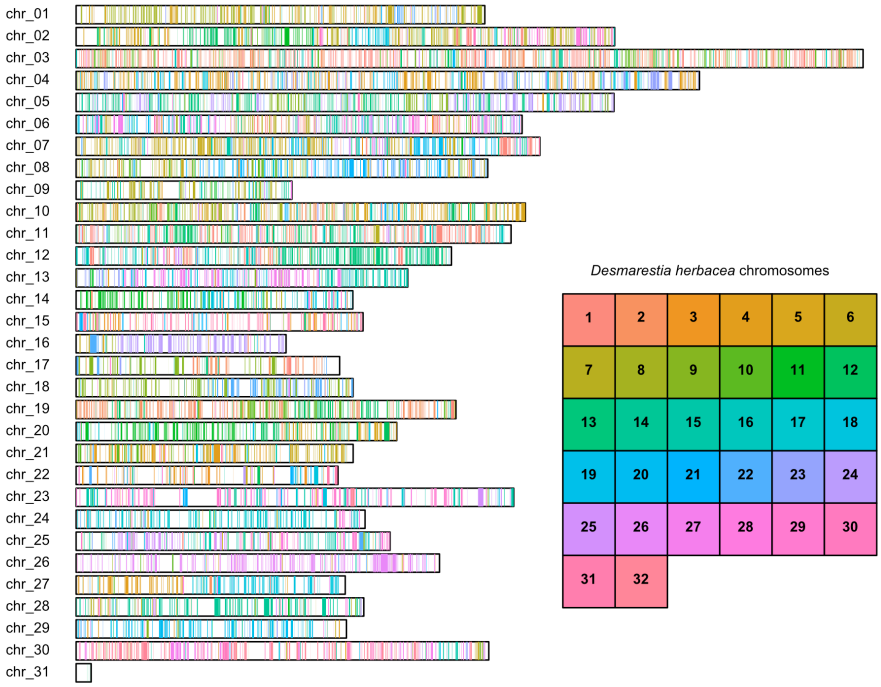



**Fig. S18**. *Schizocladia ischiensis* chromosomes painted by chromosome-of-origin for orthologous genes in *Desmarestia herbacea* (n=8,643).

## Supplemental Tables Legends

**Table S1**. Genomic data used in this study.

**Table S2**. Genome statistics (whole genome assembly + chromosome level assembly + unanchored contigs).

**Table S3**. Gene annotation statistics for species used in this study (male and monoicous data).

**Table S4**. Statistics of number of Hi-C reads for each sample and replicates from Juicer pipeline[101]

**Table S5**: Hi-C resolution used for different analysis across brown algal species

**Table S6**: Pearson correlation of linearly transformed Hi-C matrix at multiple resolutions for orthologous chromosomes between species. Rearranged chromosomes excluded from the analysis.

## References

1. Szalay, M. F., Majchrzycka, B., Jerković, I., Cavalli, G. & Ibrahim, D. M. Evolution and function of chromatin domains across the tree of life. *Nat Struct Mol Biol* **31**, 1824–1837 (2024).
2. Dekker, J. & Mirny, L. A. The chromosome folding problem and how cells solve it. *Cell* **187**, 6424–6450 (2024).
3. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, (2009).
4. Hsieh, T.-H. S. *et al.* Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **162**, 108–119 (2015).
5. Hsieh, T.-H. S., Fudenberg, G., Goloborodko, A. & Rando, O. J. Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. *Nat Methods* **13**, 1009–1011 (2016).
6. Krietenstein, N. *et al.* Ultrastructural Details of Mammalian Chromosome Architecture. *Molecular Cell* **78**, 554-565.e7 (2020).
7. Hsieh, T.-H. S. *et al.* Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Molecular Cell* **78**, 539-553.e8 (2020).
8. Sun, L. *et al.* Mapping nucleosome-resolution chromatin organization and enhancer-promoter loops in plants using Micro-C-XL. *Nat Commun* **15**, 35 (2024).
9. Lukyanchikova, V. *et al.* Anopheles mosquitoes reveal new principles of 3D genome organization in insects. *Nat Commun* **13**, 1960 (2022).
10. Feng, S. *et al.* Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in Arabidopsis. *Molecular Cell* **55**, 694–707 (2014).
11. Concia, L. *et al.* Wheat chromatin architecture is organized in genome territories and transcription factories. *Genome Biol* **21**, 104 (2020).
12. Montgomery, S. A. *et al.* Chromatin Organization in Early Land Plants Reveals an Ancestral Association between H3K27me3, Transposons, and Constitutive Heterochromatin. *Current Biology* **30**, 573-588.e7 (2020).
13. Kurbidaeva, A. *et al.* Topologically associating domains and the evolution of three-dimensional genome architecture in rice. *The Plant Journal* **122**, e70139 (2025).
14. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458–472 (2012).
15. Marín-Gual, L. *et al.* Meiotic cohesin RAD21L shapes 3D genome structure and transcription in the male germline. *Science Advances* **11**, eadv2283 (2025).
16. Bredeson, J. V. *et al.* Conserved chromatin and repetitive patterns reveal slow genome evolution in frogs. *Nature Communications* **15**, (2024).
17. Simakov, O. *et al.* Deeply conserved synteny and the evolution of metazoan chromosomes. *Sci Adv* **8**, eabi5884 (2022).
18. Varoquaux, N. *et al.* Accurate identification of centromere locations in yeast genomes using Hi-C. *Nucleic Acids Research* **43**, 5331–5339 (2015).

19. Hehmeyer, J., Spitz, F. & Marlow, H. Shifting landscapes: the role of 3D genomic organizations in gene regulatory strategies. *Current Opinion in Genetics & Development* **81**, 102064 (2023).

20. Álvarez-González, L. & Ruiz-Herrera, A. Evolution of 3D Chromatin Folding. *Annual Review of Animal Biosciences* **29**, 8 (2025).

21. Bringloe, T. T. *et al.* Phylogeny and evolution of the brown algae. *Critical Reviews in Plant Sciences* **39**, 281–321 (2020).

22. Choi, S.-W. *et al.* Ordovician origin and subsequent diversification of the brown algae. *Current Biology* **34**, 740-754.e4 (2024).

23. Coelho, S. M., Gueno, J., Lipinska, A. P., Cock, J. M. & Umen, J. G. UV chromosomes and haploid sexual systems. *Trends Plant Sci* **23**, 794–807 (2018).

24. Arun, A. *et al.* Convergent recruitment of TALE homeodomain life cycle regulators to direct sporophyte development in land plants and brown algae. *Elife* **8**, (2019).

25. Barrera-Redondo, J. *et al.* Origin and evolutionary trajectories of brown algal sex chromosomes. *Nat Ecol Evol* 1–18 (2025) doi:10.1038/s41559-025-02838-w.

26. Denoeud, F. *et al.* Evolutionary genomics of the emergence of brown algae as key components of coastal ecosystems. *Cell* **187**, 6943-6965.e39 (2024).

27. Damas, J. *et al.* Evolution of the ancestral mammalian karyotype and syntenic regions. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2209139119 (2022).

28. Batista, R. A., Wang, L., Bogaert, K. A. & Coelho, S. M. Insights into the molecular bases of multicellular development from brown algae. *Development* **151**, dev203004 (2024).

29. Umen, J. & Coelho, S. Algal sex determination and the evolution of anisogamy. *Annu. Rev. Microbiol.* **73**, 267–291 (2019).

30. Coelho, S. M. *et al.* OUROBOROS is a master regulator of the gametophyte to sporophyte life cycle transition in the brown alga *Ectocarpus*. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 11518–11523 (2011).

31. Liu, P. *et al.* 3D chromatin maps of a brown alga reveal U/V sex chromosome spatial organization. *Nat Commun* **15**, 9590 (2024).

32. Cock, J. M. *et al.* The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* **465**, 617–621 (2010).

33. Coelho, S. M. The brown seaweed Ectocarpus. *Nat Methods* **21**, 363–364 (2024).

34. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).

35. Dong, P. *et al.* 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments. *Molecular Plant* **10**, 1497–1509 (2017).

36. Karaaslan, E. S. *et al.* Marchantia TCP transcription factor activity correlates with three-dimensional chromatin structure. *Nat. Plants* **6**, 1250–1261 (2020).

37. Heesch, S. *et al.* Evolution of life cycles and reproductive traits: Insights from the brown algae. *Journal of Evolutionary Biology* **34**, 992–1009 (2021).

38. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, (2015).

39. Bianchini, G. & Sánchez-Baracaldo, P. TREEVIEWER : Flexible, modular software to visualise and manipulate phylogenetic trees. *Ecology and Evolution* **14**, e10873 (2024).

40. Dinatale, E., Craig, R. J., Martinho, C., Drost, H.-G. & Coelho, S. M. Characterization of the transposable element landscape shaping the Ectocarpus genome. *Genome Biol* **26**, 320 (2025).

41. Lovell, J. T. *et al.* GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife* **11**, (2022).

42. Yang, T. *et al.* HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* **27**, 1939–1949 (2017).

43. Serizay, J., Matthey-Doret, C., Bignaud, A., Baudry, L. & Koszul, R. Orchestrating chromosome conformation capture analysis with Bioconductor. *Nat Commun* **15**, 1072 (2024).

44. Kim, I. V. Chromatin loops are an ancestral hallmark of the animal regulatory genome. *Nature* 1–9 (2025) doi:10.1038/s41586-025-08960-w.

45. Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Systems* **3**, 99–101 (2016).

46. Montgomery, S. A. *et al.* Chromatin Organization in Early Land Plants Reveals an Ancestral Association between H3K27me3, Transposons, and Constitutive Heterochromatin. *Curr Biol* **30**, 573-588.e7 (2020).

47. Ahmed, S. *et al.* A haploid system of sex determination in the brown alga *Ectocarpus* sp. *Curr. Biol.* **24**, 1945–1957 (2014).

48. Cossard, G. G. *et al.* Selection drives convergent gene expression changes during transitions to co-sexuality in haploid sexual systems. *Nat Ecol Evol* **6**, 579–589 (2022).

49. Oluwadare, O., Zhang, Y. & Cheng, J. A maximum likelihood algorithm for reconstructing 3D structures of human chromosomes from chromosomal contact data. *BMC Genomics* **19**, 161 (2018).

50. Talbert, P. B. & Henikoff, S. What makes a centromere? *Exp Cell Res* **389**, 111895 (2020).

51. Helsen, J., Ramachandran, K., Sherlock, G. J. & Dey, G. Centromeres evolve progressively through selection at the kinetochore interface. (2025) doi:10.1101/2025.01.16.633479.

52. Bousios, A., Kakutani, T. & Henderson, I. R. Centrophilic Retrotransposons of Plant Genomes. *Annu Rev Plant Biol* **76**, 579–604 (2025).

53. Naish, M. *et al.* The genetic and epigenetic landscape of the Arabidopsis centromeres. *Science* **374**, eabi7489 (2021).

54. Chen, J. *et al.* A complete telomere-to-telomere assembly of the maize genome. *Nat Genet* **55**, 1221–1231 (2023).

55. Bi, G. *et al.* Near telomere-to-telomere genome of the model plant Physcomitrium patens. *Nat. Plants* **10**, 327–343 (2024).

56. Craig, R. J. *et al.* The Chlamydomonas Genome Project, version 6: Reference assemblies for mating-type plus and minus strains reveal extensive structural mutation in the laboratory. *Plant Cell* **35**, 644–672 (2022).

57. Liu, D. *et al.* Two CENH3 paralogs in the green alga Chlamydomonas reinhardtii have a redundantly essential function and associate with ZeppL-LINE1 elements. *Plant J* **122**, e70153 (2025).

58. Arkhipova, I. R. & Yushenova, I. A. Giant Transposons in Eukaryotes: Is Bigger Better? *Genome Biol Evol* **11**, 906–918 (2019).

59. Schubert, I. & Lysak, M. A. Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends in Genetics* **27**, 207–216 (2011).

60. Muffato, M. *et al.* Reconstruction of hundreds of reference ancestral genomes across the eukaryotic kingdom. *Nat Ecol Evol* **7**, 355–366 (2023).

61. Liu, C., Cheng, Y. J., Wang, J. W. & Weigel, D. Prominent topologically associated domains differentiate global chromatin packing in rice from Arabidopsis. *Nature Plants* **3**, 742–748 (2017).

62. Ho, J. W. K. *et al.* Comparative analysis of metazoan chromatin organization. *Nature* **512**, 449–452 (2014).

63. Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin Domains: The Unit of Chromosome Organization. *Molecular Cell* **62**, 668–680 (2016).

64. Sun, J. *et al.* Genomic and epigenomic insight into giga-chromosome architecture and adaptive evolution of royal lily (Lilium regale). *Nat Commun* **16**, 5617 (2025).

65. Wang, M. *et al.* Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nature Plants* **4**, 90–97 (2018).

66. Wen, Z. *et al.* Nucleosome wrapping states encode principles of 3D genome organization. *Nat Commun* **16**, 352 (2025).

67. Konstantinidou, P. *et al.* A comparative roadmap of PIWI-interacting RNAs across seven species reveals insights into de novo piRNA-precursor formation in mammals. *Cell Rep* **43**, 114777 (2024).

68. Akram, H. *et al.* Replication Timing Uncovers a Two-Compartment Nuclear Architecture of Interphase Euchromatin in Maize. 2025.07.11.664491 Preprint at https://doi.org/10.1101/2025.07.11.664491 (2025).

69. Simakov, O. *et al.* Deeply conserved synteny resolves early events in vertebrate evolution. *Nat Ecol Evol* **4**, 820–830 (2020).

70. Lewin, T. D., Liao, I. J.-Y. & Luo, Y.-J. Conservation of bilaterian genome structure is the exception, not the rule. *Genome Biology* **26**, 247 (2025).

71. Kim, B. Y. *et al.* Highly contiguous assemblies of 101 drosophilid genomes. *Elife* **10**, e66405 (2021).

72. Wright, C. J., Stevens, L., Mackintosh, A., Lawniczak, M. & Blaxter, M. Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera. *Nat Ecol Evol* **8**, 777–790 (2024).

73. Gonzalez de la Rosa, P. M. *et al.* A telomere-to-telomere assembly of Oscheius tipulae and the evolution of rhabditid nematode chromosomes. *G3 (Bethesda)* **11**, jkaa020 (2021).

74. Murat, F. *et al.* Understanding Brassicaceae evolution through ancestral genome reconstruction. *Genome Biol* **16**, 262 (2015).

75. Brazier, T. & Glémin, S. Diversity and determinants of recombination landscapes in flowering plants. *PLoS Genet* **18**, e1010141 (2022).

76. Waters, P. D. *et al.* Microchromosomes are building blocks of bird, reptile, and mammal chromosomes. *Proc Natl Acad Sci U S A* **118**, e2112494118 (2021).

77. Romiguier, J., Ranwez, V., Douzery, E. J. P. & Galtier, N. Contrasting GC-content dynamics across 33 mammalian genomes: Relationship with life-history traits and chromosome sizes. *Genome Res.* **20**, 1001–1009 (2010).

78. Zhu, Z., Younas, L. & Zhou, Q. Evolution and regulation of animal sex chromosomes. *Nat Rev Genet* **26**, 59–74 (2024).

79. Vigneau, J. *et al.* Rewiring of chromatin regulation underlies the evolution of brown algal multicellularity. Preprint at https://doi.org/10.1101/2025.09.16.676480 (2025).

80. Chang, C.-H. *et al.* Islands of retroelements are major components of Drosophila centromeres. *PLoS Biol* **17**, e3000241 (2019).

81. Naish, M. & Henderson, I. R. The structure, function, and evolution of plant centromeres. *Genome Res.* **34**, 161–178 (2024).

82. Neumann, P. *et al.* Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mob DNA* **2**, 4 (2011).

83. Gao, X., Hou, Y., Ebina, H., Levin, H. L. & Voytas, D. F. Chromodomains direct integration of retrotransposons to heterochromatin. *Genome Res.* **18**, 359–369 (2008).

84. Tsukahara, S. *et al.* Centrophilic retrotransposon integration via CENH3 chromatin in Arabidopsis. *Nature* **637**, 744–748 (2025).

85. Piras, F. M. *et al.* Uncoupling of satellite DNA and centromeric function in the genus Equus. *PLoS Genet* **6**, e1000845 (2010).

86. Liu, Y. *et al.* Pan-centromere reveals widespread centromere repositioning of soybean genomes. *Proceedings of the National Academy of Sciences* **120**, e2310177120 (2023).

87. Mandáková, T., Hloušková, P., Koch, M. A. & Lysak, M. A. Genome Evolution in Arabideae Was Marked by Frequent Centromere Repositioning. *Plant Cell* **32**, 650–665 (2020).

88. Rocchi, M., Archidiacono, N., Schempp, W., Capozzi, O. & Stanyon, R. Centromere repositioning in mammals. *Heredity (Edinb)* **108**, 59–67 (2012).

89. Gebert, D. *et al.* Analysis of 30 chromosome-level Drosophila genome assemblies reveals dynamic evolution of centromeric satellite repeats. *Genome Biol* **26**, 63 (2025).

90. Starr, R. C. & Zeikus, J. A. UTEX—THE CULTURE COLLECTION OF ALGAE AT THE UNIVERSITY OF TEXAS AT AUSTIN 1993 LIST OF CULTURES. *Journal of Phycology* **29**, 1–106 (1993).

91. Han, J. *et al.* An upgraded method of high-throughput chromosome conformation capture (Hi-C 3.0) in cotton (Gossypium spp.). *Frontiers in Plant Science* **14**, (2023).

92. Koren, S. *et al.* Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Res* **27**, 722–736 (2017).

93. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* **37**, 540–546 (2019).

94. Hu, J. *et al.* NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biology* **25**, 107 (2024).

95. Shafin, K. *et al.* Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. *Nat Biotechnol* **38**, 1044–1053 (2020).

96. Dudchenko, O. *et al.* De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).

97. Zeng, X. *et al.* Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. *Nat. Plants* **10**, 1184–1200 (2024).

98. Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639–1643 (2021).

99. Baril, T., Galbraith, J. & Hayward, A. Earl Grey: A Fully Automated User-Friendly Transposable Element Annotation and Analysis Pipeline. *Molecular Biology and Evolution* **41**, msae068 (2024).

100. Zhu, F. *et al.* A chromosome-level genome assembly for the Silkie chicken resolves complete sequences for key chicken metabolic, reproductive, and immunity genes. *Commun Biol* **6**, 1233 (2023).

101. Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Systems* **3**, 95–98 (2016).

102. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).

103. Open2C *et al.* Cooltools: Enabling high-resolution Hi-C analysis in Python. *PLOS Computational Biology* **20**, e1012067- (2024).

104. Knight, P. A. & Ruiz, D. A fast algorithm for matrix balancing. *IMA Journal of Numerical Analysis* **33**, 1029–1047 (2013).

105. Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).

106. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

107. Goubert, C. *et al.* Correction: A beginner's guide to manual curation of transposable elements. *Mob DNA* **13**, 15 (2022).

108. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772–780 (2013).

109. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

110. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530–1534 (2020).

111. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238 (2019).

112. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

113. Kruse, K., Hug, C. B. & Vaquerizas, J. M. FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. *Genome Biology* **21**, 303 (2020).

114. Sweeten, A. P., Schatz, M. C. & Phillippy, A. M. ModDotPlot—rapid and interactive visualization of tandem repeats. *Bioinformatics* **40**, btae493 (2024).