Repeated co-option of HMG-box genes for sex determination in brown algae and animals

Luthringer, Rémy; Raphalen, Morgane; Guerra, Carla; Colin, Sébastien; Martinho, Claudia; Zheng, Min

Published in:
Science

DOI:
10.1126/science.adk5466

Publication date:
2024

Document Version
Peer reviewed version

Author Accepted Manuscript: Luthringer, Rémy et al. "Repeated co-option of HMG-box genes for sex determination in brown algae and animals". Science. 2024. 383(6689). https://doi.org/10.1126/science.adk5466

Title: Repeated co-option of HMG-box genes for sex determination in brown algae and animals

Authors: Rémy Luthringer[1†], Morgane Raphalen[1†], Carla Guerra[1], Sébastien Colin[1], Claudia Martinho[1], Min Zheng[1], Masakazu Hoshino[1,2], Yacine Badis[3], Agnieszka P Lipinska[1], Fabian B. Haas[1], Josué Barrera-Redondo[1], Vikram Alva[4], Susana M. Coelho[1*]

Affiliations:

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

[2]Research Center for Inland Seas, Kobe University, Japan

[3]Roscoff Biological Station, CNRS-Sorbonne University, Place Georges Teissier, Roscoff, 29680 France.

[4]Department of Protein Evolution, Max Planck Institute for Biology Tübingen, 72076 Tübingen, Germany.

[†]Contributed equally

[*]Corresponding author. Email: susana.coelho@tuebingen.mpg.de.

Abstract: In many eukaryotes, genetic sex determination is not governed by XX/XY or ZW/ZZ systems but by a specialized region on the poorly studied U (female) or V (male) sex chromosomes. Previous studies have hinted at the existence of a dominant male-sex factor on the V chromosome in brown algae, a group of multicellular eukaryotes distantly related to animals and plants. The nature of this factor has remained elusive. Here, we demonstrate that an HMG-box gene acts as the male-determining factor in brown algae, mirroring the role HMG-box genes play in sex determination in animals. Over a billion-year evolutionary timeline, these lineages have independently co-opted the HMG-box for male determination, representing a remarkable paradigm for evolution's ability to recurrently use the same genetic 'toolkit' to accomplish similar tasks.

One sentence summary: Animals and brown seaweeds have independently converged on the same solution for determining male sex.

Main text

Three major sex chromosome systems have been identified across the diverse organisms comprising the Eukaryotic Tree of Life. The XX/XY and ZZ/ZW systems are found in diploid sexual organisms, including animals and land plants. The U/V sex system exists in eukaryotes with haploid-diploid life cycles, including the bryophytes and brown, red and green algae (*1*). In the XX/XY system, males bear XY while females carry XX chromosomes; in the ZZ/ZW system, males possess ZZ chromosomes and females ZW; and in the unique U/V system, females possess a U chromosome and males a V chromosome (**Fig. 1A**). Sex-determining regions on the Y, W, V or U chromosomes typically harbor the factor(s) responsible for triggering the male versus female sex differentiation network. Although XY and ZW sex chromosomes have been characterised in several organisms, only a few master sex-determining factors are currently known. These include mammalian *SRY* (*2*), avian *DMRT1* (*3*), fish DMY (*4*, *5*), and feminizing factors like the -KTS splice variant of WT1 in mouse (*6*), *popARR17* in poplar (*7*) and *MeGl* in *Diospyros lotus* (*8*). Knowledge regarding master sex-determining factors in U/V system is even more limited: only one female-determining factor has been identified. This factor, present on the U chromosome of *Marchantia polymorpha,* is plant-specific (*9*). Note that in fungi, mating-type determination is often controlled by a specialized region of the genome (the *MAT* locus), that may have properties similar to sex chromosomes (e.g., (*1*, *10*, *11*).

Despite the apparent diversity of eukaryotic sex-determination mechanisms, members of conserved gene families have been linked to sex determination in animals (*12*, *13*). For example, the *doublesex/mab-3 related* (*Dmrt*) family of transcription factors shape sexual dimorphism in organisms as diverse as mammals, insects, and nematodes (*14*). A key member of this family, the DM domain-containing *DMRT1* gene, is part of the mammalian sex-determination cascade, and it initiates sex-determination in other vertebrate lineages (*15*, *16*). Similarly, the high mobility group (HMG)-box domain, a small domain that binds the minor grove of DNA, is encoded by various genes involved in sex determination: the *SRY* gene that initiates male sex determination in therian mammals, various *SOX* genes involved in vertebrate sex determination, and mating-type (*MAT*) genes which are crucial for mating type specification in fungi (*17–21*). Furthermore, the HMG-box domain is remarkably widespread in other eukaryotic proteins, particularly in non-histone components of chromatin and transcription factors, which often contain multiple copies that bind DNA non-specifically (*21*). Underpinning its widespread occurrence, the most recent version of the InterPro database (v97.0; entry IPR009071) lists a significant presence of HMG-box proteins across eukaryotic species: 149 in humans, 133 in mice, 113 in zebrafish, and 8 in baker's yeast. Unlike chromatin-associated proteins, HMG-box proteins linked to sex determination typically contain a single copy of the HMG-box domain, flanked by intrinsically disordered regions, and bind DNA specifically. In the SRY protein of therian mammals, the HMG-box domain acts as a transcription factor that regulates other transcription factors, most importantly SOX9, eventually resulting in the development of testes (*2*, *22*, *23*). In addition to the specific binding of the SRY HMG-box domain to promoter sites, its ability to induce significant conformational changes in

DNA is thought to facilitate the recruitment of other factors, enabling the precise regulation of sex determination in mammals (*24*).Likewise, the HMG-box domain in the fungal MAT proteins acts as a transcription regulator, controlling the expression of genes crucial for mating type specification and sexual reproduction (*19*).

A prominent example of the under-explored U/V sex chromosome system is found in brown algae, a group of photosynthetic multicellular organisms that thrive in coastal areas globally. As members of the Stramenopiles supergroup, they have evolved independently from animals and land plants for over a billion years (7), becoming the third most developmentally complex lineage on the planet (*25*). Notably, kelps, a subgroup of brown algae, display spectacular morphologies, forming underwater forests with immense ecological significance (*26*). Most brown algal species exhibit separate sexes, determined by U/V sex chromosomes (*1*, *27*) (**Fig. 1A**). Sex is determined at meiosis (not at fertilization as in diploid XX/XY and ZW/ZZ systems), depending on whether daughter cells inherit a U or V chromosome, and then develop as multicellular females or males, respectively. The level of sexual dimorphism is variable across the different brown algae species, spanning from near-isogamy (males and female gametes are very similar) to strongly oogamous (males and females gametes are strongly dimorphic) (*28*). Genetic and genomic analyses using the model brown alga *Ectocarpus sp.* (*25*, *29*) and the giant kelp *Macrocystis pyrifera* (*30*) suggest that the V sex chromosome carries a masculinising factor, whereas the U chromosome likely harbors loci required for a complete female developmental program (*31*). Moreover, algae possessing both U and V chromosomes develop as males, supporting the presence of a dominant V-linked masculinising locus (*32*, *33*). However, the identity of this sex-determining factor has remained elusive. Here, we report the identification of the brown alga master male-determining gene thereby revealing a remarkable example of convergent evolution.

Identification of a candidate male-determining gene

To identify sex-determining genes, we examined genomes spanning the brown algal phylogeny and representing over 200 million years of evolution (*34*). A single gene was V-limited in all studied species (*35*). Intriguingly, this gene encodes a putative transcription factor featuring two HMG-box domains and a nuclear localization signal, sharing similarities with mammalian *SRY* (*2*) and fungal *MAT* genes (*11*, *19*, *36*) (**Fig. 1B, 1C**). Henceforth, we refer to this gene as *HMG-sex*. Using Colabfold (*37*, *38*) we predicted the three-dimensional structure of the *Ectocarpus sp.7* HMG-sex, revealing a canonical first HMG-box domain and a divergent second one with a large insertion (**Fig. 1C**, **Fig. S1**). Additionally, HMG-sex contains an N-terminal domain with a TANGO2-like fold, characteristic of the TANGO2 family of eukaryotic and prokaryotic proteins whose functions are still largely unknown, although the human TANGO2 protein, a member of this family, has recently been implicated in lipid homeostasis (*39*). Note that the TANGO2-like domain of HMG-sex exhibits no discernible sequence similarity to TANGO2 domains of other proteins, suggesting it may have a distinct biological activity in brown algae.

An RNA-seq approach was used to investigate the expression pattern of *HMG-sex* during the *Ectocarpus* life cycle. This analysis uncovered strong upregulation in the male mature gametophyte stage of development when male sex is presumably determined (**Fig. 1D**). Although these findings suggest a role for this gene in male sex in brown algae, direct evidence is necessary to confirm that *HMG-sex* is indeed required for male sex determination.

To test this hypothesis, we generated knockout (KO) HMG-sex lines using the CRISPR/Cas system (*40*, *41*). We focused on the *HMG-sex* locus of the brown algal model *Ectocarpus sp.* (Ec-13_0001750), for which genetic and genomic tools are available, and whose development is well characterized (*25*, *40*, *42–44*). Introduction of CRISPR/Cas protein and guide RNAs targeting *HMG-sex* resulted in three independent mutant lines (*hmg1-3;* **Fig. 1E**, **Table S1-S4**). These strains contain an additional mutation at the *APT* locus that allows selection of mutants (*40*) (**Table S2**). *hmg1* has a +34 bp insertion, *hmg2* has a -4 bp deletion, and *hmg3* has an in-frame deletion (-6 bp), (**Table S1-S4**, **Fig. 1E, Fig. S2**). Mutations in *hmg1* and *hmg2* lead to a premature stop codon and a significantly disrupted predicted protein. In contrast, *hmg3* carries a -2 amino acid mutation and the structure of the predicted protein is expected to remain unaffected (**Fig. 1E, Figs. S2-S3**). Consequently, *hmg3* represents a convenient additional control for specifically studying the effects of *HMG-sex* disruption.

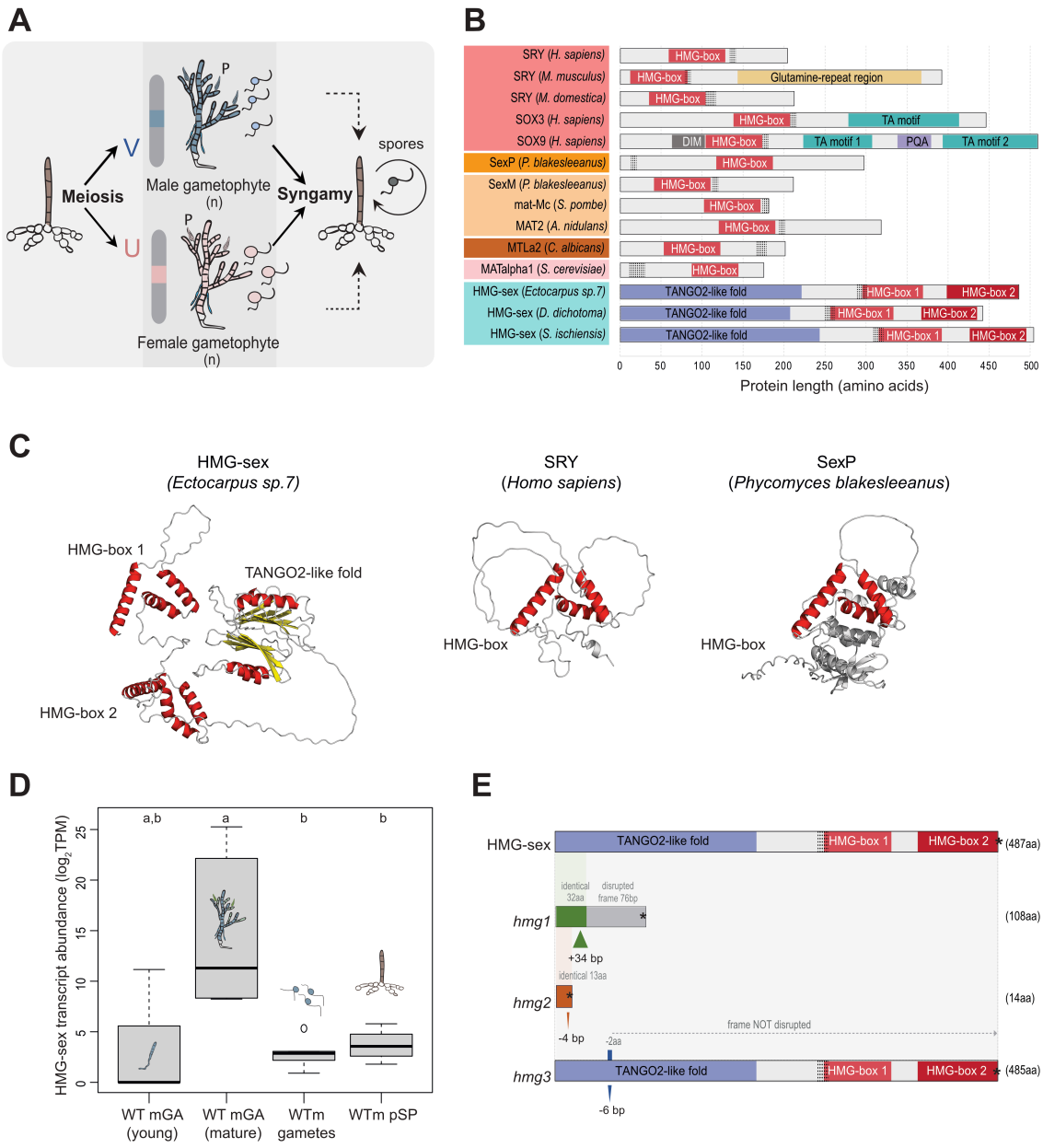

Fig. 1. An HMG-box protein coding gene located on the V-sex chromosome is a candidate master male-determining gene in brown algae. (**A**) Schematic view of the life cycle of *Ectocarpus*. Meiosis occurs in the sporophyte. The daughter cells that inherit a V sex chromosome develop into males, whereas cells that inherit a U sex chromosome develop into females. Gametes are produced in plurilocular gametangia (P), released into the seawater, and syngamy produces a sporophyte. Alternative pathways (dashed line) via parthenogenesis may occur (*45*, *46*) and gametes develop directly into partheno-sporophytes. **(B)** Protein domain composition of HMG-box domain-containing proteins involved in sex determination in eukaryotes. The known domains are highlighted for each protein, and the nuclear localization signals are shown as dotted patterns within the protein. **(C)** AlphaFold2-predicted structures of HMG-sex in *Ectocarpus* alongside SRY in *Homo sapiens* and SexP in *Phycomyces blakesleeanus*. α-helices are colored red and β-strands yellow. **(D)** Expression level (log_2TPM) of *HMG-sex* during the life cycle of *Ectocarpus*. WT mGA: wild-type male gametophyte; WTm pSP: wild-type male partheno-sporophyte. Different letters above the plot indicate significant differences (Wilcoxon test). **(E)** Scheme of WT HMG-sex protein and predicted proteins of each of the CRISPR-generated *hmg* mutants. Asterisk indicates a stop codon; aa: amino-acid; bp: base pair.

HMG-sex is required for the production of functional male gametes in *Ectocarpus*

We examined the phenotype of the *hmg1*, *hmg2* and *hmg3* mutants compared to control (WT and *apt*) male and female lines during various stages of the life cycle. We found no evidence for morphological feminization in the gametophytes of the three *hmg* mutants, and, overall, no distinctive modification in the morphology of the *hmg* partheno-sporophytes compared to wild-type (WT) males and *apt* lines (**Fig. 2A**). This result is consistent with the notion that *Ectocarpus* WT male and female gametophytes display modest sexual dimorphism (*28*, *47*, *48*), with sexual differences mainly restricted to the gamete stage (*28*, *49*). We therefore tested whether the zoids produced by the *hmg* KO mutant lines behaved as fully functional male gametes.

Gamete fusion in *Ectocarpus* follows a stepwise process that begins with the attraction of male gametes by the pheromone ectocarpene produced by settled female gametes (*50*), and then proceeds with gamete recognition and cell-cell fusion (*51*, *52*). We examined the mating behaviour of *apt* and *hmg1-3* mutant lines in response to female gametes, employing high-speed video microscopy. In the absence of females, all gametes exhibit a linear swimming pattern or move in large circles (*53*) and we observed no differences in this behaviour in the mutant lines compared to the WT (**Fig. S4**). Thus, mutations at the *HMG-sex* locus do not affect the swimming capability of the zoids.

In the presence of settled female gametes, the behaviour of WT male gametes undergoes a notable change. When they sense a decrease in the sex pheromone concentration gradient, their posterior flagellum engages in rapid, unilateral beating (*53*), causing reorientation towards the source of pheromone production. This results in a distinctive narrow circular trajectory of the gametes (*53*) leading to the formation of typical 'clusters' as illustrated in **Fig. 2B** (see also **Fig. S5; Movie 1, 2)**. In contrast, when *hmg1* and *hmg2* zoids were confronted with settled female gametes, their swimming pattern remained unaltered, indicating they are unable to sense the pheromone (**Fig. 2B, Fig. S5, Movie 3, 4**). To confirm that the unresponsiveness of the mutant zoids was not due to decreased pheromone production by the females, we mixed WT male gametes with the same settled female gametes. Immediately, the characteristic clustering behaviour of WT male gametes was observed, indicating that the female gametes were fully functional (**Fig. S6**). Collectively, these observations suggest that whilst *hmg1* and *hmg2* mutant zoids exhibit normal free-swimming pattern, they are incapable of responding to settled female gametes, and therefore do not behave as fully functional males. Interestingly, *hmg3* (which carries a silent mutation at the *HMG* locus) behaved like WT male gametes (**Fig. 2B, Fig. S5, Movie 5**), confirming that the lack of response to the pheromone was caused by the full disruption of the HMG-sex protein.

We evaluated fertilization success by quantifying the number of zygotes produced in controlled crosses. A substantial number of zygotes were produced in crosses involving both WT male and females, whereas no zygotes were observed when *hmg1* and *hmg2* zoids were mixed with female gametes (**Fig. 2C-D**, **Table S5**). Conversely, crosses involving *apt* and *hmg3* mutants resulted in a zygote count comparable to that in the WT controls.

The inability of the *hmg1* and *hmg2* mutants to recognize and fuse with female gametes extended beyond their lack of attraction to the pheromone. Even when an abundance of mutant zoids was introduced, ensuring close proximity to female gametes, we observed no instances of gamete fusion (**Fig. S7**), suggesting that mutant zoids do not 'recognize' a female gamete even in close contact, and are unable to engage in membrane-membrane fusion events. These observations underscore that a functional HMG-sex protein is necessary for sensing the pheromone and for recognizing and fusing with gametes of the opposite sex. Taken together, HMG-sex is crucial for the manifestation of functional male characteristics in *Ectocarpus*.

The observed absence of fusion between *hmg* KO mutants and female gametes raises the intriguing possibility that *hmg* mutants may have undergone partial or even complete conversion to females. However, when we cultivated *hmg* KO zoids in isolation, we did not observe clear attraction clusters nor zygote formation (**Fig. S4**). Furthermore, when *hmg* gametes were crossed with WT male gametes, we found neither zygotes nor evidence of attraction (**Fig. S8A).** This observation indicates that in the absence of a functional HMG-sex protein, zoids produced by mutant gametophytes become fully asexual. Whilst *Ectocarpus* gametes and asexual spores are morphologically indistinguishable ((*54*); see scheme in **Fig. 1A**), they display different swimming times before settlement. Remarkably, settlement and swimming behaviour of *hmg1* and *hmg2* (but not *hmg3*) zoids indeed resembled that of asexual spores (**Fig. 2E, Fig. S8B**), and not that of male or female gametes. The absence of a functional HMG-sex, therefore, leads to the production of 'demasculinized', asexual zoids in *Ectocarpus*, but does not induce sex reversal into female.

Importantly, note that a complete sex reversion is highly unlikely in the U/V system of brown algae given the absence of a U chromosome in males and the requirement of the U-specific region for the full expression of the female program in these organisms (*30*). This distinct feature sets it apart from XX/XY diploid systems, where deletion of the master male-determining gene leads to phenotypic feminization (*20*). This is also unlike other haploid systems such as the Volvox system, where KO of *VcMID* (the RWP-RK domain transcription factor determinant of sperm and egg development) results in the production of functional eggs or self-fertile hermaphrodites (*55*). This disparity may be attributed to the more complex multicellular development of *Ectocarpus* compared with Volvox.

It is thus possible that the lack of full sex reversal in the *Ectocarpus hmg-sex* KOs is due to the absence of the U chromosome in these individuals. To test this hypothesis, we investigated whether *HMG-sex* KO in a genetic background containing the U chromosome, would allow the U-triggered female program to be expressed. We exploited diploid gametophytes (*32*, *56*) containing both the U and V chromosome that have been shown to be phenotypically male (*32*). Two CRIPSR/Cas12 mutants were examined (**Table S3**). At fertility, *UVhmg-sex* mutants released (diploid) zoids. When *UVhmg-sex* mutant zoids were confronted with WT male gametes, a large number of (triploid) zygotes were observed (**Fig. S9, Table S6**). We therefore conclude that the presence of both the U and V chromosome in the absence of a functional HMG-sex leads to sex-

reversal, likely by de-repression of the female (U-triggered) program. These observations imply that HMG-sex is a master regulator of the male sex determination in *Ectocarpus*.

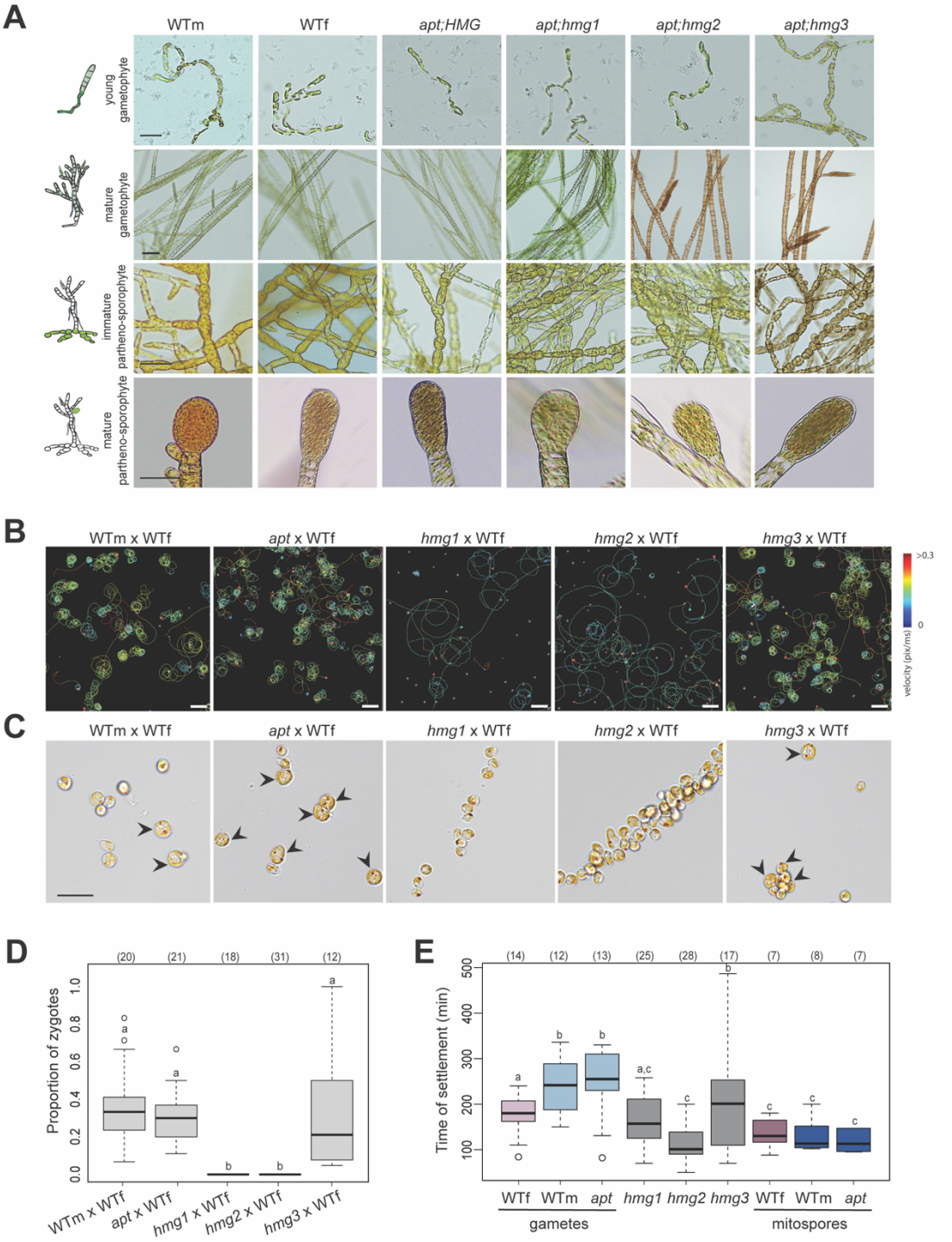

Fig. 2. Phenotype of *Ectocarpus sp.7 hmg* mutants compared with WT. (**A**) Morphological phenotype of WT and *hmg* mutants during key stages of the *Ectocarpus* life cycle, schematized to the left of the photos. Bar: 40 µm. (**B**) A 15-sec-long 2D trajectory showing the differences in swimming patterns of *Ectocarpus* WT and mutant male zoids when in contact with WT female settled gametes. See also **Fig. S5**. Bar: 50 µm. (**C**) Germlings of *Ectocarpus* $_{24h}$ after a cross. Diploid zygotes are indicated with arrow

heads. Zygotes were scored by the presence of typical two eyespots, rapid cell wall formation and large size. In contrast, no zygotes are produced when KO *hmg1* and *hmg2* zoids are confronted with WT female gametes. Bar: 10 µm. **(D)** Proportion of zygotes obtained after confronting female WT gametes with different WT male and mutant strains. Between 241 and 724 germlings were scored in n replicates (n in brackets). Different letters above the plots indicate significant differences (Wilcoxon test, $p<0.05$). See also **Table S5**. **(E)** Male, female gametes and spores of *Ectocarpus* have different timing of settlement after release from the gametangia. Different letters above the plot represent significant differences ($p<0.05$; Wilcoxon test). See also **Table S7.**

Absence of a functional HMG-sex leads to partial transcriptome feminisation and de-masculinization

To identify pathways downstream of HMG-sex and further characterize the role of this factor in sexual differentiation, we analysed the transcriptomes of *hmg* mutants and contrasted them with WT, *hmg3,* and *apt* background lines. Because the expression of *HMG-sex* was highest during the mature gametophyte stage (see **Fig. 1D**), we focused on this developmental stage. Note that although *Ectocarpus* gametophyte morphological sexual dimorphism is almost non-existent, male and female gametophytes express distinct transcriptomes (*47*). We used DEseq2 (*57*) to define candidate genes involved in sexual differentiation (sex-biased genes, SBGs, i.e., genes showing differential expression in WT males versus WT females). Overall, transcriptomic patterns of SBG in mature gametophytes of all mutant samples clustered together (**Fig. S10**). We then focused on a subset of 278 SBGs that are uniquely expressed in the presence of a functional HMG-sex (i.e., we removed SBGs that are also differentially expressed in an *apt* background), reasoning that these 278 genes are potential downstream effectors of HMG-sex (**Fig. S11**, **Table S10**). Whilst, as expected, transcriptomic patterns of the 278 genes in *hmg3* and *apt* resemble those of WT males, *hmg1* and *hmg2* expression profiles cluster with female WT, suggesting feminisation of their expression (**Fig. 3A-B**). Therefore, despite the lack of morphological effects of *HMG-sex* disruption in gametophytes (due to the absence of overall sexual dimorphism at this stage), *HMG-sex* disruption leads to gametophytic de-masculinization and feminization of a subset of SBGs. Notably, these candidate effector genes are enriched in functions related to microtubule process, cell differentiation and developmental processes (**Fig. S12**).

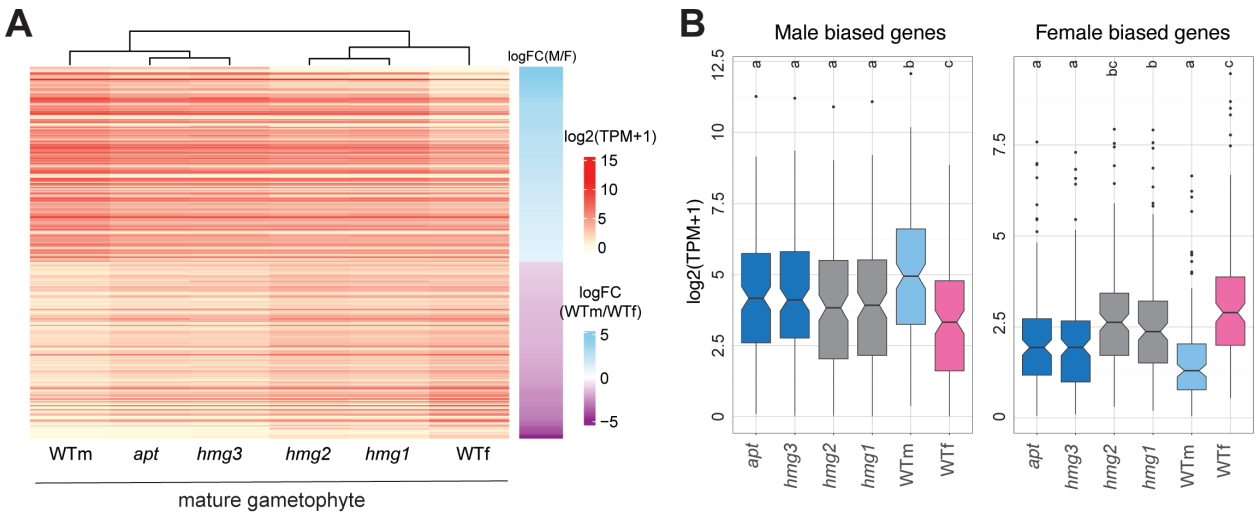

Figure 3. Transcriptome of *Ectocarpus* sp. 7 HMG-sex KO gametophytes is partially feminized and demasculinized. **(A)** Hierarchical clustering of gene expression patterns in wild-type and mutant samples at the mature gametophyte stage of development. Heatmap was built using 287 candidate HMG-effector SBGs (see **Table S8, S9**). **(B)** Transcript abundance [in log2(TPM+1)] of the 278 sex-biased genes in WT and mutant mature gametophytes. Different letters above the plots represent significant differences in expression levels (Wilcoxon test).

HMG-sex is also required for male sex determination in kelps

HMG-sex is a V-linked gene across all examined brown algae with separate sexes (*35*) suggesting a conserved function in sex determination. In order to test this hypothesis, we generated HMG-sex KO mutants in the kelp *Laminaria digitata*, which diverged from *Ectocarpus* >100 Mya (*34*). Sexual dimorphism in kelps is substantially more conspicuous than in *Ectocarpus* (*58*), providing an opportunity to investigate the role of HMG-sex in an organism with strong gametophyte sexual dimorphism. Two CRISPR/Cas LdHMG-sex mutant lines (Ld*hmg1*, Ld*hmg2*) were examined, both in an *apt* background (**Table S3, S11**). Morphological characterization of these gametophytes revealed that they are both strongly feminized (**Fig. 4A-B**). At maturity, Ld*hmg* lines did not release flagellated mobile zoids, as WT males. Instead, these lines only produced large immobile cells resembling eggs. When male gametes were mixed with these egg-like cells, no zygotes could be observed, suggesting they are not functional eggs. Remarkably, these egg-like cells were capable of parthenogenesis, which is a female-specific trait in kelps (**Fig. 4C**). Similar to *Ectocarpus*, lack of a functional HMG-sex in a kelp leads to loss of maleness but not to a full sex reversal in the absence of a U chromosome (**Fig. 4D**). Contrary to *Ectocarpus*, and consistent with the morphological sexual dimorphism in kelp gametophytes, mutant gametophytes were strongly feminized at the morphological level. Interestingly, a similar feminized phenotype was associated with silencing of HMG-sex in genetically male haploid variant lines of another kelp species (*Macrocystis pyrifera*) (*30*). Thus, HMG-sex functions as a sex-determining factor in kelps, and the partial switch to the female program in the absence of a functional HMG-sex suggests that this gene acts as a master transcription factor likely regulating the sex differentiation pathway.

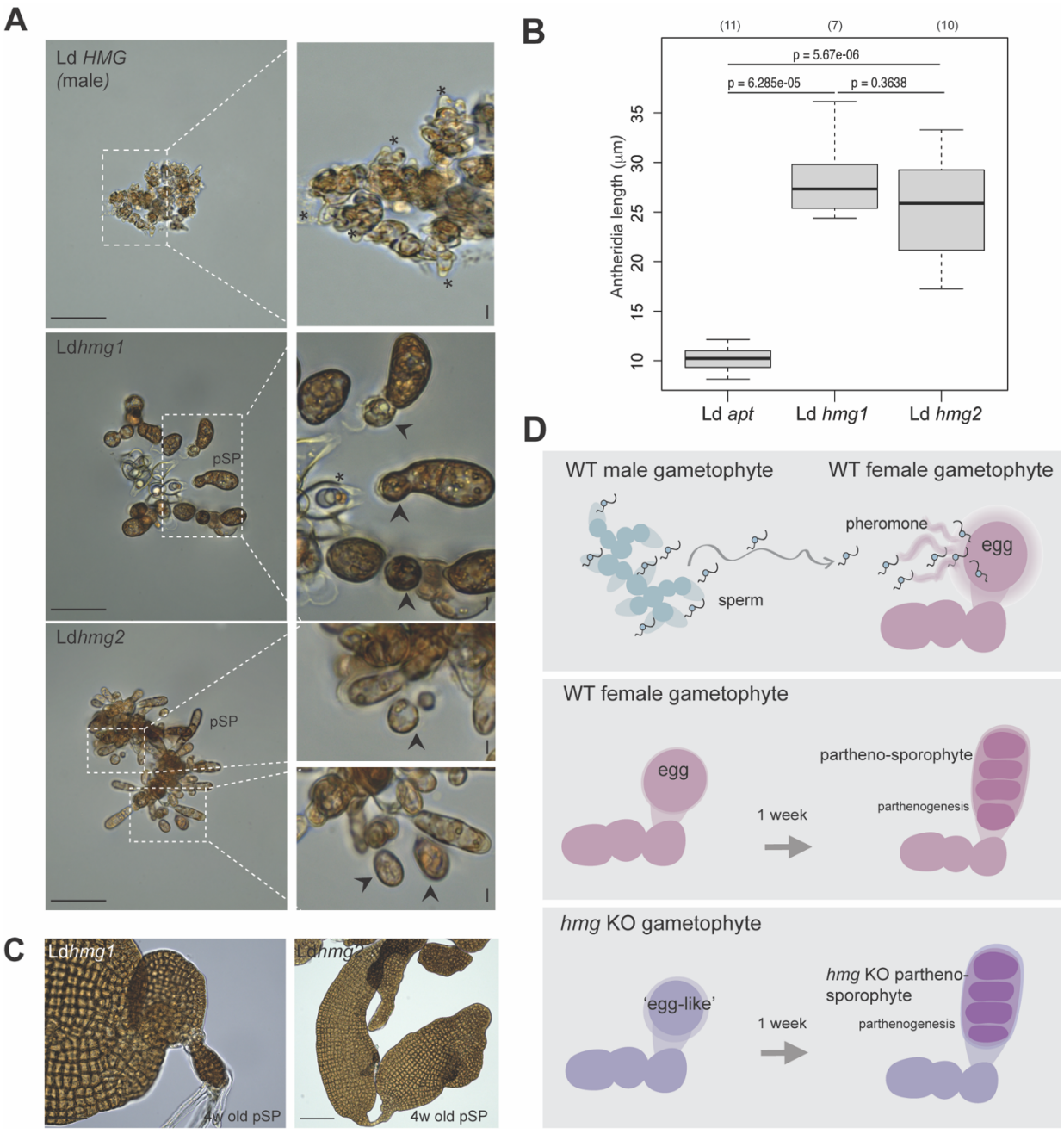

Figure 4. **HMG-sex is required for male sex determination in the kelp *L. digitata*. (A)** Representative images of mature gametophytes of control male (Ld*apt*) and Ld *HMG-sex* mutants (Ld*hmg1* and Ld*hmg2*). Note the pear-shaped gametangia (asterisks) where gametes are produced, which are significantly enlarged in *HMG-sex* mutant lines. Arrow-heads highlight egg-like structures developing parthenogenetically. Bar: 10 µm. (**B**) Mean gametangia lengths (in µm) in *apt* lines (control) and Ld*hmg-sex* mutants. Statistical analysis was performed in R (Wilcoxon test, p-values indicated above the plots). The number of replicates is presented in brackets. **(C)** Partheno-sporophytes (pSP) developing from egg-like structures in Ld*hmg1* and Ld*hmg2* mutants. Bar: 50 µm. (**D**) Schematic view of development in WT and mutants. In WT males, antheridia produce biflagellate sperm cells, which are released in the media and swim towards the female eggs in response to pheromone production (upper panel). In the absence of males, WT females can reproduce by parthenogenesis, producing partheno-sporophytes (middle panel). In Ld*hmg* mutants, 'egg-like' cells with no visible flagella are produced in enlarged gametangia. Ld*hmg* gametophytes are sterile, but the egg-like structures develop by parthenogenesis similarly to WT females (lower panel).

HMG-box genes have independently evolved sex-determining functions across eukaryotic kingdoms

To investigate the evolutionary origins of HMG-sex and its relationship with the HMG-box domain-containing mammalian SRY and fungal MAT proteins, we retrieved homologs of HMG-sex, SOX3, SRY, MAT, and other widespread HMG superfamily proteins, such as the FACT complex protein SSRP1 and non-histone chromatin-associated HMGB family proteins (e.g., HMGB1), from the UniProt database (*59*). We expanded our search for homologs of HMG-sex to publicly available brown algal genomes and transcriptomes, given the limited representation of brown algal proteins in UniProt. We then employed CLANS (*60*, *61*) to cluster the pooled sequences based on their all-against-all pairwise sequence similarities, with an E-value cut-off of 1e-12. CLANS uses the Fruchterman-Reingold force-directed layout algorithm, creating a multidimensional virtual space where protein sequences, represented as point masses, attract or repel each other based on pairwise similarity strength (*62*). This interaction leads to evolutionarily related sequences clustering in similar areas of the map. In the resulting cluster map (**Fig. 5A**), SRY formed a tight cluster with various SOX proteins, with its highest similarity to SOX3, as previously reported (*63*). While SOX proteins are ubiquitous in metazoans, SRY is confined to therian mammals and is thought to have evolved from SOX3 through a duplication event. Two distinct clusters, composed of MAT-associated proteins from the Ascomycota phylum and basal phyla Zoopagomycota and Mucoromycota (including SexP and SexM proteins of *Phycomyces blakesleeanus*), are closely linked to the SOX cluster. The close proximity of MAT and SOX proteins on the map suggests a shared origin from an ancestral protein at the root of opisthokonts. Several other well-characterized fungal MAT-associated proteins, such as MATalpha1 of *Saccharomyces cerevisiae* (Mating-type cluster 2), form distinct clusters that are not linked to the SOX or the aforementioned fungal MAT clusters at the cut-off chosen for clustering, suggesting they may represent divergent proteins. Furthermore, a cluster comprising members of the transcription Factor TCF/LEF family (InterPro IPR024940) involved in the Wnt signaling cascade, such as transcription factor 7, lymphoid enhancer-binding factor 1, and pangolin, is also tightly connected to the SOX cluster (**Fig. 5A**).

In the cluster map, HMG-sex proteins from seven distinct brown algal species, exhibiting over 35% pairwise sequence identities, form a cluster that is clearly separated from the SOX and MAT clusters. Instead, the HMG-sex cluster is linked to three distinct but tightly connected clusters, comprising the FACT complex proteins SSRP1 and Nhp6, and the HMGB family proteins (**Fig. S1**), all of which contain representatives from a diverse range of eukaryotes. Notably, while SSRP1 and Nhp6 each contain one HMG-box domain, HMGB proteins typically possess two consecutive HMG-box domains like HMG-sex. Furthermore, in addition to HMG-sex, *Ectocarpus sp.* contains 12 HMG-box domain-containing proteins, with the highest sequence similarity of HMG-sex observed with proteins from the HMGB family. This suggests that HMG-sex may have evolved from the duplication and subsequent diversification of a gene encoding for an HMGB protein. However, the alternative hypothesis that it arose from the duplication of a *SSRP1-/Nhp6-*

like gene cannot be fully excluded. In the cluster map, two metazoan-specific clusters of transcription factors, the TOX and nucleolar transcription factor 1 clusters, are also connected to the chromatin-associated FACT and HMGB clusters, suggesting that novel transcription factors may have arisen independently on multiple occasions from chromatin-associated HMG-box domains during the course of eukaryotic evolution. To obtain further hints about the origin of HMG-sex, we searched the EukProt database and transcriptomes of closely related sister groups of brown algae. Curiously, we identified an ortholog of HMG-sex, possessing all three domains and exhibiting a pairwise identity of ~30%, in *Schizocladia ischiensis*. Its role in sex determination in this organism is unclear, as the life cycle cannot be completed under laboratory conditions. This search also found matches to hypothetical proteins in species of the classes Chrysophyceae (e.g., *Chlorochromonas danica*), Xanthophyceae (*Vaucheria litorea*), and Pelagophyceae (*Aureoumbra lagunensis*) (**Fig. 5B**). However, while some of these proteins possess two HMG domains like HMG-sex, they do not possess an N-terminal domain of the TANGO2-like fold, suggesting that they are not orthologs of HMG-sex. In sum, our analysis indicates that the sex-determining role of HMG-sex likely emerged in the common ancestor of brown algae, independent of opisthokonts, possibly through the duplication of an HMGB family gene. Concurrently, our data support the notion that HMG-box genes in animals and fungi, which are involved in sex determination, have a shared evolutionary origin.

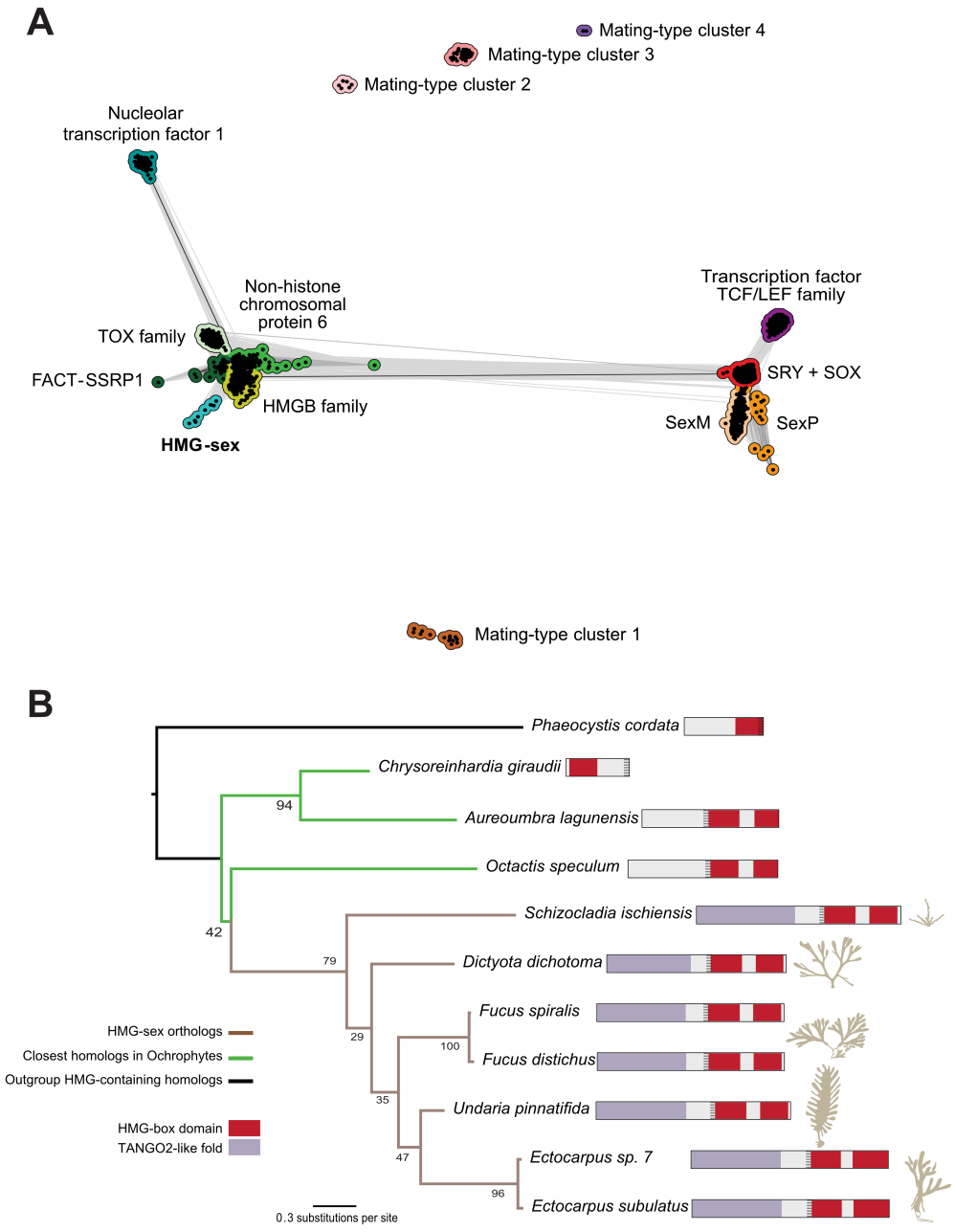

Figure 5. **HMG-box containing proteins are repeatedly co-opted as sex-determining factors throughout the tree of life (A)** Cluster map in two-dimensional space representing the evolutionary relationship of different sex-determining genes with HMG domains. We compiled homologs of HMG-box domain-containing proteins related to sex determination and others representative of the HMG-box domain superfamily, clustering them based on all-against-all pairwise sequence similarity strengths. Each dot on the map represents a sequence, with sequences from the same group color-coded identically. The darkness of the connecting lines indicates the significance of sequence similarities, with darker lines representing higher significance. **(B)** Maximum likelihood tree with bootstrap values displaying the evolutionary relationship and structure of HMG-sex across brown algae and other HMG-box domain-containing proteins in Ochrophyta. The nuclear localization signals are shown as dotted patterns within the protein.

Conclusion

Our experiments uncover a male-limited *Ectocarpus* gene *HMG-sex* (Ec-13_0001750) that encodes an HMG-box domain transcription factor. Loss of this gene is associated with loss of male-specific characteristics, including the ability to sense female pheromone, recognize female gametes and gamete fusion to produce zygotes. The gametophyte transcriptome of these mutants also shows partial de-masculinization and feminization. Without a functional *HMG-sex*, *Ectocarpus* gametophytes produce 'demasculinized', asexual zoids. The kelp *HMG-sex* ortholog also has a male-determining role, indicating the function of this protein has been conserved for more than 100 Mya. However, the absence of HMG-sex alone is insufficient to feminize the mutants, likely due to the lack of a U-specific region that is required to fully activate the female developmental program. Indeed, the deletion of HMG-sex in male individuals carrying both U and V chromosomes leads to sex reversal, further supporting the idea that this gene is the master male-determining gene in *Ectocarpus* and demonstrating that female is not the 'default' program in this organism. We named this gene MIN (Male-INducer) after Min, an ancient Egyptian god associated with male fertility and virility. Our investigation provides evidence that MIN, similar to SRY in mammals, is required for viable male gamete formation and male-sex determination in brown algae.

HMG-box genes play a role in sex and mating-type determination in the distantly related Opisthokonta group, which encompasses animals and fungi. Opistokonts and brown algae diverged over a billion years ago, and the sexes in brown algae, animals and fungi arose independently. Thus, the similarities observed in master sex-determining factors between brown algae and opistokonts could be attributed to either a shared ancestry or convergent evolution (*64*). Our analysis supports the independent co-option hypothesis: even though HMG-box genes are involved in sex determination across lineages, the specific sex-determining genes have evolved independently. Thus, the HMG-box domain, likely owing to its ability to induce large structural changes in DNA, enabling precise control of gene regulation, has been co-opted multiple times to play a role in sex determination across eukaryotic kingdoms, providing a fascinating example of convergent evolution over a billion-year timescale.

Material and Methods

Biological material

Table S1 describes the strains used. Ec32 (*Ectocarpus* sp. 7) is the reference genome strain (*65*). We used the male *Ectocarpus* species 7 strain Ec32, for which a reference genome is available (*65*, *66*). Strain Ec32 was previously referred to as *E. siliculosus* (*67*). However, a recent study (*68*) indicates that it belongs to a distinct, at present unnamed, species, which is referred to provisionally as *Ectocarpus* species 7 (*Ectocarpus sp.7*). For simplicity, we use the term '*Ectocarpus*' here. The kelp *Laminaria digitata* material was collected in Santec, France and cultured in sterilized natural seawater enriched with half-strength Provasoli solution (*69*). All strains were cultured at 14°C with a light: dark cycle of 12 h : 12 h (30 µmol m^{-2} s^{-1}), and daylight-

type LEDs (adapted from (*70*, *71*)). All manipulations were performed under sterile conditions in a laminar flow hood.

Generation of mutants by CRISPR/Cas

We followed the protocol described in (*40*, *41*). Our experimental system uses a single reporter gene (*UpAPT*) encoding adenine phosphoribosyl transferase (APT) and generates a resistant phenotype when algae are cultured in a medium with the toxic compound 2-fluoroadenine. Therefore, all *hmg* mutants isolated also have a mutation at the *APT* locus (**Table S1**). Note that *apt* mutant lines have a noticeable change in their global transcriptomes so our transcriptomic comparisons were established using *apt* lines as controls. Mutations were detected by PCR amplification and Sanger sequencing. Guide RNAs and PCR primers are described in **Table S2.**

Fertilisation success and mating behavior

Reproductive success was assessed by measuring the production of zygotes in controlled experimental crosses (*72*). In brief, we mixed the same amount of male and female tissue in a suspending drop, and the proportion of gametes that succeeded in fusing (i.e., form a zygote) was scored (**Table S5**). Three independent parental crosses were performed. In each cross, several fields were observed and the presence of germlings (unfused gametes or zygotes) was recorded. Between 241 and 724 germlings were counted for each cross.

Mating behavior was recorded by high-speed video microscopy. Before recording, female gametes or zoids to be tested as potential pheromones producers, were allowed to settle. Then, zoids were added to test their 'maleness'. In the case of lack of attraction and fusion, and in order to test the "receptiveness" of female gametes, a suspension of WT male gametes was added approximately 20 minutes after the previous zoids suspension. All cross experiments, including the video recording, were performed at 14°C.

In order to further study the behavior of the zoids, we investigated the speed at which zoids settle after release from reproductive structures. We scored the timing of zoid settlement for the different lines studied (WT versus mutants) using an inverted microscope (Zeiss, Axiovert AX10).

Motility measurements

After allowing the female gametes to attach to the bottom surface of the imaging dish, the media was removed and replaced with a suspension of male gametes. In order to characterize the behavior of male gametes, 15 sec videos at 70 frames per second (14,29 msec per frame) were acquired with an inverted microscope (Zeiss, Axiovert AX10) equipped with a global shutter cmos camera (Imaging Source, DFK37AUX250), under transmitted light illumination, with a 10x objective LD APlan(NA0.25), at a the focal plane of the female gametes (circa 10 µm of depth of field).

The tracking of gamete motion was carried out with Fiji (*73*) and the plugin Trackmate (v7.11.1) (*74*). Before undertaking the tracking, images of gametes were segmented according to the following operations within Fiji: conversion of the video into a sequence of still images, image

smoothing with a median filter (radius 2), autothresholding (Isodata method), binarization, holes filling within the object of interests in binary images, and watershed operation for splitting touching objects. Briefly, detected spots were filtered according to their size so that spots below 15 pixels area and above 150 pixels area (white color in the figure) were removed from the analysis. The Kalman tracker algorithm was used to track the spots (allowing a search radius of 30 pixels (circa twice the diameter of the segmented gametes) and a gap of 15 frames and the objects with trajectories displaying a maximum distance travelled (this feature reports the distance to the furthest point of the track, with respect to the first spot in time of the track) below 25 pixels are considered as attached and non-motile. They appear within a grey circle on the figures whereas other spots are encircled according to the duration of their track with a jet colormap ranging values from blue (0 sec) to red (15 sec). Tracking conflicts caused by crossing gametes and out of focus elements were not included as they do not impact our analysis. The instantaneous velocity of the gametes is encoded in the figures with a jet colormap ranging values from blue (0.0 pixel/msec) to red (above 0.3 pixel/msec). The videos with the tracking overlay are presented as supplementary movies. Quantitative analysis of the zoid swimming path was performed as in (*51*, *53*). In brief, the distribution of the Menger curvature value K was calculated along the motion trajectory of the zoids for each tracking experiment with the different mating combinations. Menger curvature K of three points in 2-dimensional Euclidean space is the reciprocal of the radius of the unique circumcircle that passes through the three points (*75*). The distribution of K was measured for each position P(x,y) at any timepoint T(i) along tracks. The tracks were manually curated to ensure accuracy of the cell tracking and spurious tracks were removed. For each mating experiment, a minimum of n=40 tracks were analysed. See also **Table S12**.

DNA extraction

Genomic DNA was extracted with the OmniPep™ for plant kit (G-Biosciences, Cat. No. 786-397) according to the manufacturer's protocol with some modifications. Briefly, the tissue was ground to a fine powder in liquid nitrogen. 50-100 mg of finely ground tissue was quickly transferred to a microcentrifuge tube containing 500 μl of Lysis Buffer and carefully mixed by inverting. Then, 5 μl of Proteinase K was added to the solution and incubated at 60 °C for 1-2 hours with periodic inversion every 15 minutes. After that, 200 μl of chloroform was added and a 10 minutes centrifugation at 14000 g was performed. The upper phase was removed to a clean microcentrifuge tube and 50 μl of DNA Stripping Solution was added. After a 10 minutes incubation at 60 °C, 100 μl of Precipitation Solution was added and it was incubated on ice for 15 minutes followed by a 10 minutes centrifugation at 14000 g. The supernatant was transferred to a clean tube and 500 μl of isopropanol was added. After pelleting the genomic DNA was centrifuged at 14000 g for 10 minutes, and an ethanol 75% wash was performed. Ethanol was removed and the pellet dried at room temperature, 50 μl of TE buffer was added to the pellet and clean DNA was incubated at 60 °C for 30 minutes. Finally, 1 μl of LongLife™ RNase was added for every 100 μl of TE buffer and it was incubated 30 minutes at 60 °C. The samples were stored at -80 °C until further use.

Screen for off-target mutations

The genome of the *hmg1* mutant strain was sequenced on an Illumina NextSeq 2000 platform, generating 20Gb of sequence data, corresponding to 13.4 million 150 bp paired-end reads (accession numbers in **Table S1, S2**).The sequenced reads were analyzed according to the variant calling pipeline by (*76*). Software updates were accomplished for GATK version 4.2.6.1 (*77*) and gmap-gsnap (version 2021-12-17) (*78*). The reads were mapped to the Ec32 reference genome (version v2, available at Orcae (*79*). Potential off-target sites were predicted using Crispor (*80*) (**Table S3**) and searches for mutations at potential off-target sites were carried out using both the table of variants and by manual visualization of the genomic regions using the Jbrowse2 genome viewer (*81*).

RNA extraction and transcriptomic analysis

RNAseq was used to characterize the transcriptome of the mutant lines compared to similar stages in WT male and female lines. Biological material was grown in the conditions described above and samples were frozen at specific stages of development. Material was flash frozen, RNA extracted using an adapted Qiagen RNAeasy procedure (as in (13)) and TruSeq RNA Library Prep Kit v2 was used to sequence the transcriptomes in an Illumina NextSeq 2000 platform.

RNA-seq reads from each library were used to quantify gene expression with kallisto v.0.44.0 (*82*) using 31-base-pair-long k-mers, 1000 bootstraps and *Ectocarpus* sp. 7 transcriptome as reference (https://bioinformatics.psb.ugent.be/orcae/overview/EctsiV2). Transcript abundances were then summed within genes using the tximport package (*83*) to obtain the expression level for each gene in transcripts per million (TPM).

Estimates of sex-biased gene expression and pairwise differential expression between the mutants and the wild-type were obtained using read count matrices as input for the DESeq2 package (*57*) in R v.4.3.1. P-values were corrected for multiple testing using Benjamini and Hochberg's algorithm in DESeq2, applying an adjusted P-value cut-off of 0.05 for differential expression analysis. In addition, only genes with a minimum of 2-fold change (FC) expression level between sexes were retained.

Structural model of *Ectocarpus* HMG-sex

Structure predictions were conducted using ColabFold (*37*), which integrates MMseqs2 for sequence similarity search and AlphaFold2 (*38*) for protein structure prediction. Given that publicly available sequence databases have limited HMG-sex sequences, many of which are incomplete, we opted for a custom-built multiple sequence alignment (MSA) as input for ColabFold. Initially, we constructed an MSA containing the HMG-sex sequence from *Ectocarpus sp. 7* and manually corrected HMG-sex sequences from various brown algal species, including *Fucus distichus*, *Fucus spiralis*, *Dictyota dichotoma*, and *Undaria pinnatifida* (*66*, *84–86*). The alignment was computed using MAFFT (*87*). Subsequently, this seed MSA was used to build a larger MSA by running three iterations of HHblits (*88*) against the Uniclust30 database (*37*). We

employed default settings when running the ColabFold pipeline using our custom MSA. Following the prediction, the resulting models were ranked according to their pTM scores and the highest-ranking model was selected for further analysis (**Fig. S13**). The prediction was carried out on the high-performance computer "Raven", operated by the Max-Planck Computing & Data Facility in Garching, Munich, Germany. The AlphaFold2 structures of human SRY and *P. blakesleeanus* SexP shown in **Fig. 1** were downloaded from the UniProt database. To assess the domain and fold composition of HMG-sex, we ran HHpred (*89*) searches against the PDB70 and ECOD70 databases, and Foldseek (*61*, *90*) searches against the AlphaFold/UniProt50 v4 database, both with default settings.

Evolution of HMG-sex in the brown algae

We extracted brown algal HMG-box domain-containing proteins from publicly available genomic resources (*66*, *84–86*). We searched for putative orthologs of HMG-sex using OrthoFinder (*91*). We manually corrected the gene models of the HMG-sex orthologs using GenomeView (*92*) to verify that they were complete to assess its structural conservation in different species. We retrieved complete HMG-sex protein sequences for *Ectocarpus* sp.7, *Ectocarpus subulatus*, *Undaria pinnatifida*, *Fucus distichus*, *Fucus serratus*, *Dictyota dichotoma* and *Schizocladia ischiensis* (*66*, *84–86*).

To identify protein sequences for cluster analysis, we queried the UniProt database (*93*) for homologs of various HMG-box domain-containing proteins. Specifically, we focused on HMG-sex proteins from *Ectocarpus* sp. 7 and *Dictyota dichotoma*, the HMG-sex-like protein from *Schizocladia ischiensis*, and the human sex-determining region Y protein (SRY; UniProt ID: Q05066). Additionally, we searched for homologs of HMG-box domain-containing proteins implicated in the mating-type determination in fungi, including *Aspergillus nidulans* (UniProt IDs: AAP92161, G5EAT5), *Saccharomyces cerevisiae* (P0CY06), *Schizosaccharomyces pombe* (P0CY17), *Neurospora crassa* (P19392, P36981, Q10116), *Candida albicans* (Q71U11, Q9UW19), and *Phycomyces blakesleeanus* (B0F2H1, A0A167RE73). Beyond proteins involved in sex and mating-type determination, we also considered other HMG-box domain-containing proteins such as yeast non-histone chromosomal protein 6A (P11632) and human proteins SSRP1 (Q08945), HMGB1 (P09429), SOX3 (P41225), transcription factor 7 (P36402), TOX (O94900), and nucleolar transcription factor 1 (P17480) (**Supplemental Dataset 1**). For the sequence similarity searches against the UniProt database, we used the HMG domains of the aforementioned proteins, predicted using InterPro (*94*), as seed sequences. Sequence similarity searches were conducted using BLAST(*93*), employing an E-value threshold of 10^{-16} and setting the 'max_target_seqs' parameter to 20,000. The full-length sequences of the resulting matches were aggregated (sequences flagged as ‘Fragment’ were removed) and subsequently filtered using MMseqs2 (*95*) to retain sequences with a maximum pairwise identity of 70% at a length coverage of at least 70%. This yielded a total of 3,491 candidate sequences. These filtered sequences were then subjected to clustering analysis using CLANS (*60*, *61*). Since these proteins belong to different HMG-box domain-containing families that are highly divergent, we opted for a clustering

analysis over a phylogenetic analysis to explore their evolutionary relationships. We based the clustering on all-against-all pairwise P-values, calculated using BLAST. The clustering was performed until equilibrium was reached in a 2D space by applying a P-value cutoff of 1e-12 and using the default settings in CLANS.

To further extend our search for closely-related homologs of HMG-sex proteins, we also examined the EukProt database (*96*). We downloaded version 3 of EukProt and built a BLAST-searchable database using the 'makeblastdb' command with default settings. BLAST searches were then conducted, also using default settings, with the HMG-sex protein sequence from *Ectocarpus* sp. 7 as the query. The highest bitscore hits from sister groups of brown algae were aligned with the HMG-sex orthologs using MAFFT (*87*) to reconstruct a maximum likelihood tree using RAxML (*97*). We used a protein CAT approximation and the corrected Akaike Information Criterion from RAxML to select the best empirical substitution matrix for the dataset. 100 bootstrap permutations were performed to assess the support values at each node in the tree. The HMG-box domain-containing protein of *Phaeocystis cordata* was used to root the tree. We predicted the presence of nuclear localization signals using DeepLoc 2.0 (*98*).

References

1. S. M. Coelho, J. Gueno, A. P. Lipinska, J. M. Cock, J. G. Umen, UV chromosomes and haploid sexual systems. *Trends Plant Sci* **23**, 794–807 (2018).

2. P. Koopman, J. Gubbay, N. Vivian, P. Goodfellow, R. Lovell-Badge, Male development of chromosomally female mice transgenic for Sry. *Nature* **351**, 117–121 (1991).

3. L. Beukeboom, N. Perrin, *The Evolution of Sex Determination* (Oxford University Press, 2015).

4. M. Matsuda, Y. Nagahama, A. Shinomiya, T. Sato, C. Matsuda, T. Kobayashi, C. E. Morrey, N. Shibata, S. Asakawa, N. Shimizu, H. Hori, S. Hamaguchi, M. Sakaizumi, DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* **417**, 559–563 (2002).

5. I. Nanda, M. Kondo, U. Hornung, S. Asakawa, C. Winkler, A. Shimizu, Z. Shan, T. Haaf, N. Shimizu, A. Shima, A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, Oryzias latipes. *Proceedings of the National Academy of Sciences* **99**, 11778–11783 (2002).

6. E. P. Gregoire, M.-C. De Cian, R. Migale, A. Perea-Gomez, S. Schaub, N. Bellido-Carreras, I. Stévant, C. Mayère, Y. Neirijnck, A. Loubat, P. Rivaud, M. L. Sopena, S. Lachambre, M. M. Linssen, P. Hohenstein, R. Lovell-Badge, S. Nef, F. Chalmel, A. Schedl, M.-C. Chaboissier, The −*KTS* splice variant of WT1 is essential for ovarian determination in mice. *Science* **382**, 600–606 (2023).

7. N. A. Müller, B. Kersten, A. P. Leite Montalvão, N. Mähler, C. Bernhardsson, K. Bräutigam, Z. Carracedo Lorenzo, H. Hoenicka, V. Kumar, M. Mader, B. Pakull, K. M. Robinson, M. Sabatti, C. Vettori, P. K. Ingvarsson, Q. Cronk, N. R. Street, M. Fladung, A single gene underlies the dynamic evolution of poplar sex determination. *Nat. Plants* **6**, 630–637 (2020).

8. T. Akagi, I. M. Henry, R. Tao, L. Comai, A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science* **346**, 646–650 (2014).

9. M. Iwasaki, T. Kajiwara, Y. Yasui, Y. Yoshitake, M. Miyazaki, S. Kawamura, N. Suetsugu, R. Nishihama, S. Yamaoka, D. Wanke, K. Hashimoto, K. Kuchitsu, S. A. Montgomery, S. Singh, Y. Tanizawa, M. Yagura, T. Mochizuki, M. Sakamoto, Y. Nakamura, C. Liu, F. Berger, K. T. Yamato, J. L. Bowman, T. Kohchi, Identification of the sex-determining factor in the liverwort Marchantia polymorpha reveals unique evolution of sex chromosomes in a haploid system. *Current Biology* **31**, 5522-5532.e7 (2021).

10. S. Branco, H. Badouin, R. C. Rodriguez de la Vega, J. Gouzy, F. Carpentier, G. Aguileta, S. Siguenza, J.-T. Brandenburg, M. A. Coelho, M. E. Hood, T. Giraud, Evolutionary strata on

young mating-type chromosomes despite the lack of sexual antagonism. *Proc Natl Acad Sci U S A* **114**, 7067–7072 (2017).

11. C. M. Hull, M.-J. Boily, J. Heitman, Sex-specific homeodomain proteins Sxi1alpha and Sxi2a coordinately regulate sexual development in Cryptococcus neoformans. *Eukaryot Cell* **4**, 526–535 (2005).

12. A. Herpin, M. Schartl, Plasticity of gene-regulatory networks controlling sex determination: of masters, slaves, usual suspects, newcomers, and usurpators. *EMBO Rep* **16**, 1260–1274 (2015).

13. C. S. Raymond, C. E. Shamu, M. M. Shen, K. J. Seifert, B. Hirsch, J. Hodgkin, D. Zarkower, Evidence for evolutionary conservation of sex-determining genes. *Nature* **391**, 691–695 (1998).

14. A. Kopp, Dmrt genes in the development and evolution of sexual dimorphism. *Trends in Genetics* **28**, 175–184 (2012).

15. S. Chen, G. Zhang, C. Shao, Q. Huang, G. Liu, P. Zhang, W. Song, N. An, D. Chalopin, J.-N. Volff, Y. Hong, Q. Li, Z. Sha, H. Zhou, M. Xie, Q. Yu, Y. Liu, H. Xiang, N. Wang, K. Wu, C. Yang, Q. Zhou, X. Liao, L. Yang, Q. Hu, J. Zhang, L. Meng, L. Jin, Y. Tian, J. Lian, J. Yang, G. Miao, S. Liu, Z. Liang, F. Yan, Y. Li, B. Sun, H. Zhang, J. Zhang, Y. Zhu, M. Du, Y. Zhao, M. Schartl, Q. Tang, J. Wang, Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nat Genet* **46**, 253–260 (2014).

16. D. Zarkower, M. W. Murphy, DMRT1: An Ancient Sexual Regulator Required for Human Gonadogenesis. *Sex Dev* **16**, 112–125 (2022).

17. W. Li, T. D. Sullivan, E. Walton, A. F. Averette, S. Sakthikumar, C. A. Cuomo, B. S. Klein, J. Heitman, Identification of the mating-type (MAT) locus that controls sexual reproduction of Blastomyces dermatitidis. *Eukaryot Cell* **12**, 109–117 (2013).

18. A. Idnurm, F. J. Walton, A. Floyd, J. Heitman, Identification of the sex genes in an early diverged fungus. *Nature* **451**, 193–196 (2008).

19. J. Ait Benkhali, E. Coppin, S. Brun, L. Peraza-Reyes, T. Martin, C. Dixelius, N. Lazar, H. van Tilbeurgh, R. Debuchy, A Network of HMG-box Transcription Factors Regulates Sexual Cycle in the Fungus Podospora anserina. *PLoS Genet* **9**, e1003642 (2013).

20. S. Kurtz, A. Lucas-Hahn, B. Schlegelberger, G. Göhring, H. Niemann, T. C. Mettenleiter, B. Petersen, Knockout of the HMG domain of the porcine SRY gene causes sex reversal in gene-edited pigs. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2008743118 (2021).

21. M. Štros, D. Launholt, K. D. Grasser, The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cell. Mol. Life Sci.* **64**, 2590–2606 (2007).

22. C. Larney, T. L. Bailey, P. Koopman, Switching on sex: transcriptional regulation of the testis-determining gene *Sry*. *Development* **141**, 2195–2205 (2014).

23. I. Georg, S. Bagheri-Fam, K. C. Knower, P. Wieacker, G. Scherer, V. R. Harley, Mutations of the SRY-Responsive Enhancer of *SOX9* Are Uncommon in XY Gonadal Dysgenesis. *Sex Dev* **4**, 321–325 (2010).

24. M. H. Werner, J. R. Huth, A. M. Gronenborn, G. Marius Clore, Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell* **81**, 705–714 (1995).

25. S. M. Coelho, A. F. Peters, D. Müller, J. M. Cock, Ectocarpus: an evo-devo model for the brown algae. *EvoDevo* **11**, 19 (2020).

26. G. Conroy, Why Earth's giant kelp forests are worth $500 billion a year. *Nature*, d41586-023-01307–3 (2023).

27. S. M. Coelho, L. Mignerot, J. M. Cock, Origin and evolution of sex-determination systems in the brown algae. *New Phytologist* **10.1111/nph.15694** (2019).

28. R. Luthringer, A. Cormier, A. F. Peters, J. M. Cock, S. M. Coelho, Sexual dimorphism in the brown algae. *Perspectives in Phycology* **1**, 11–25 (2015).

29. J. Cock, O. Godfroy, M. Strittmatter, D. Scornet, T. Uji, G. Farnham, A. Peters, S. Coelho, "Emergence of *Ectocarpus* as a model system to study the evolution of complex multicellularity in the brown algae" (2015), pp. 153–162.

30. D. G. Müller, E. Gaschet, O. Godfroy, J. Gueno, G. Cossard, M. Kunert, A. F. Peters, R. Westermeier, W. Boland, J. M. Cock, A. P. Lipinska, S. M. Coelho, A partially sex-reversed giant kelp sheds light into the mechanisms of sexual differentiation in a UV sexual system. *New Phytologist* **232**, 252–263 (2021).

31. D. G. Müller, E. Gaschet, O. Godfroy, J. Gueno, G. Cossard, M. Kunert, A. F. Peters, R. Westermeier, W. Boland, J. M. Cock, A. P. Lipinska, S. M. Coelho, A partially sex-reversed giant kelp sheds light into the mechanisms of sexual differentiation in a UV sexual system. *bioRxiv*, 2021.02.28.433149 (2021).

32. S. M. Coelho, O. Godfroy, A. Arun, G. Le Corguillé, A. F. Peters, J. M. Cock, 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).

33. S. Ahmed, J. M. Cock, E. Pessia, R. Luthringer, A. Cormier, M. Robuchon, L. Sterck, A. F. Peters, S. M. Dittami, E. Corre, M. Valero, J.-M. Aury, D. Roze, Y. Van de Peer, J. Bothwell, G. A. B. Marais, S. M. Coelho, A haploid system of sex determination in the brown alga *Ectocarpus* sp. *Curr. Biol.* **24**, 1945–1957 (2014).

34. T. T. Bringloe, S. Starko, R. M. Wade, C. Vieira, H. Kawai, O. D. Clerck, J. M. Cock, S. M. Coelho, C. Destombe, M. Valero, J. Neiva, G. A. Pearson, S. Faugeron, E. A. Serrão, H.

Verbruggen, Phylogeny and evolution of the brown algae. *Critical Reviews in Plant Sciences* **39**, 281–321 (2020).

35. A. P. Lipinska, N. R. T. Toda, S. Heesch, A. F. Peters, J. M. Cock, S. M. Coelho, Multiple gene movements into and out of haploid sex chromosomes. *Genome Biology* **18**, 104 (2017).

36. D. Dooijes, M. van de Wetering, L. Knippels, H. Clevers, The Schizosaccharomyces pombe mating-type gene mat-Mc encodes a sequence-specific DNA-binding high mobility group box protein. *Journal of Biological Chemistry* **268**, 24813–24817 (1993).

37. M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679–682 (2022).

38. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

39. A. L. Lujan, O. Foresti, C. Sugden, N. Brouwers, A. M. Farre, A. Vignoli, M. Azamian, A. Turner, J. Wojnacki, V. Malhotra, Defects in lipid homeostasis reflect the function of TANGO2 in phospholipid and neutral lipid metabolism. *eLife* **12**, e85345 (2023).

40. Y. Badis, D. Scornet, M. Harada, C. Caillard, O. Godfroy, M. Raphalen, C. M. M. Gachon, S. M. Coelho, T. Motomura, C. Nagasato, J. M. Cock, Targeted CRISPR-Cas9-based gene knockouts in the model brown alga Ectocarpus. *New Phytol* **231**, 2077–2091 (2021).

41. K. Ichihara, T. Yamazaki, S. Kawano, Genome editing using a DNA -free clustered regularly interspaced short palindromic repeats -CAS9 system in green seaweed *ULVA PROLIFERA* . *Phycological Research* **70**, 50–56 (2022).

42. O. Godfroy, T. Uji, C. Nagasato, A. P. Lipinska, D. Scornet, A. F. Peters, K. Avia, S. Colin, L. Mignerot, T. Motomura, J. M. Cock, S. M. Coelho, DISTAG/TBCCd1 Is Required for Basal Cell Fate Determination in *Ectocarpus*. *Plant Cell* **29**, 3102–3122 (2017).

43. O. Godfroy, M. Zheng, H. Yao, A. Henschen, A. F. Peters, D. Scornet, S. Colin, P. Ronchi, K. Hipp, C. Nagasato, T. Motomura, J. M. Cock, S. M. Coelho, The *baseless* mutant links protein phosphatase 2A with basal cell identity in the brown alga *Ectocarpus*. *Development* **150**, dev201283 (2023).

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

45. L. Mignerot, K. Avia, R. Luthringer, A. P. Lipinska, A. F. Peters, J. M. Cock, S. M. Coelho, A key role for sex chromosomes in the regulation of parthenogenesis in the brown alga *Ectocarpus*. *PLoS Genet* **15**, e1008211 (2019).

46. J. H. Bothwell, D. Marie, A. F. Peters, J. M. Cock, S. M. Coelho, Role of endoreduplication and apomeiosis during parthenogenetic reproduction in the model brown alga *Ectocarpus*. *New Phytol* **188**, 111–121 (2010).

47. Lipinska, A. Cormier, R. Luthringer, A. F. Peters, E. Corre, C. M. M. Gachon, J. M. Cock, S. M. Coelho, Sexual dimorphism and the evolution of sex-biased gene expression in the brown alga *Ectocarpus*. *Mol Biol Evol* **32**, 1581–1597 (2015).

48. G. G. Cossard, O. Godfroy, Z. Nehr, C. Cruaud, J. M. Cock, A. P. Lipinska, S. M. Coelho, Selection drives convergent gene expression changes during transitions to co-sexuality in haploid sexual systems. *Nat Ecol Evol* **6**, 579–589 (2022).

49. J. Gueno, M. Borg, S. Bourdareau, G. Cossard, O. Godfroy, A. Lipinska, L. Tirichine, J. M. Cock, S. M. Coelho, Chromatin landscape associated with sexual differentiation in a UV sex determination system. *Nucleic Acids Res* **50**, 3307–3322 (2022).

50. I. Maier, Brown algal pheromones. *Progress in Plycological Research* **11** (1995).

51. A. Geller, D. G. Müller, Analysis of the flagellar beat pattern of male *Ectocarpus siliculosus* gametes (Phaeophyta) in relation to chemotactic stimulation by female cells. *J. Exp. Biol.* **92**, 53- (1981).

52. I. Maier, C. Schmid, An immunofluorescence study on lectin binding sites in gametes of Ectocarpus siliculosus (Ectocarpales, Phaeophyceae). *Phycological Research* **43**, 33–42 (2006).

53. N. Kinoshita, K. Shiba, K. Inaba, G. Fu, C. Nagasato, T. Motomura, Flagellar waveforms of gametes in the brown alga *Ectocarpus siliculosus*. *European Journal of Phycology* **51**, 139–148 (2016).

54. B. Charrier, S. M. Coelho, A. Le Bail, T. Tonon, G. Michel, P. Potin, B. Kloareg, C. Boyen, A. F. Peters, J. M. Cock, Development and physiology of the brown alga *Ectocarpus siliculosus*: two centuries of research. *New Phytologist* **177**, 319–332 (2008).

55. S. Geng, P. De Hoff, J. G. Umen, Evolution of sexes from an ancestral mating-type specification pathway. *PLoS Biol* **12**, e1001904 (2014).

56. S. M. Coelho, O. Godfroy, A. Arun, G. Le Corguillé, A. F. Peters, J. M. Cock, Genetic regulation of life cycle transitions in the brown alga *Ectocarpus*. *Plant Signal Behav* **6**, 1858–1860 (2011).

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

58. S. Heesch, M. Serrano-Serrano, J. Barrera-Redondo, R. Luthringer, A. F. Peters, C. Destombe, J. M. Cock, M. Valero, D. Roze, N. Salamin, S. M. Coelho, Evolution of life cycles and reproductive traits: insights from the brown algae. *Journal of Evolutionary Biology* **n/a** (2021).

59. The UniProt Consortium, A. Bateman, M.-J. Martin, S. Orchard, M. Magrane, S. Ahmad, E. Alpi, E. H. Bowler-Barnett, R. Britto, H. Bye-A-Jee, A. Cukura, P. Denny, T. Dogan, T. Ebenezer, J. Fan, P. Garmiri, L. J. Da Costa Gonzales, E. Hatton-Ellis, A. Hussein, A. Ignatchenko, G. Insana, R. Ishtiaq, V. Joshi, D. Jyothi, S. Kandasaamy, A. Lock, A. Luciani, M. Lugaric, J. Luo, Y. Lussi, A. MacDougall, F. Madeira, M. Mahmoudy, A. Mishra, K. Moulang, A. Nightingale, S. Pundir, G. Qi, S. Raj, P. Raposo, D. L. Rice, R. Saidi, R. Santos, E. Speretta, J. Stephenson, P. Totoo, E. Turner, N. Tyagi, P. Vasudev, K. Warner, X. Watkins, R. Zaru, H. Zellner, A. J. Bridge, L. Aimo, G. Argoud-Puy, A. H. Auchincloss, K. B. Axelsen, P. Bansal, D. Baratin, T. M. Batista Neto, M.-C. Blatter, J. T. Bolleman, E. Boutet, L. Breuza, B. C. Gil, C. Casals-Casas, K. C. Echioukh, E. Coudert, B. Cuche, E. De Castro, A. Estreicher, M. L. Famiglietti, M. Feuermann, E. Gasteiger, P. Gaudet, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz, C. Hulo, N. Hyka-Nouspikel, F. Jungo, A. Kerhornou, P. Le Mercier, D. Lieberherr, P. Masson, A. Morgat, V. Muthukrishnan, S. Paesano, I. Pedruzzi, S. Pilbout, L. Pourcel, S. Poux, M. Pozzato, M. Pruess, N. Redaschi, C. Rivoire, C. J. A. Sigrist, K. Sonesson, S. Sundaram, C. H. Wu, C. N. Arighi, L. Arminski, C. Chen, Y. Chen, H. Huang, K. Laiho, P. McGarvey, D. A. Natale, K. Ross, C. R. Vinayaka, Q. Wang, Y. Wang, J. Zhang, UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* **51**, D523–D531 (2023).

60. T. Frickey, A. Lupas, CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* **20**, 3702–3704 (2004).

61. L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A. N. Lupas, V. Alva, A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology* **430**, 2237–2243 (2018).

62. T. M. J. Fruchterman, E. M. Reingold, Graph drawing by force-directed placement. *Softw Pract Exp* **21**, 1129–1164 (1991).

63. M. Stevanovlć, R. Lovell-Badge, J. Collignon, P. N. Goodfellow, *SOX3* is an X-linked gene related to *SRY*. *Hum Mol Genet* **2**, 2013–2018 (1993).

64. J. A. Graves, C. L. Peichel, Are homologies in vertebrate sex determination due to shared ancestry or to limited options? *Genome Biol* **11**, 205 (2010).

65. J. M. Cock, L. Sterck, P. Rouzé, D. Scornet, A. E. Allen, G. Amoutzias, V. Anthouard, F. Artiguenave, J.-M. Aury, J. H. Badger, B. Beszteri, K. Billiau, E. Bonnet, J. H. Bothwell, C. Bowler, C. Boyen, C. Brownlee, C. J. Carrano, B. Charrier, G. Y. Cho, S. M. Coelho, J. Collén, E. Corre, C. Da Silva, L. Delage, N. Delaroque, S. M. Dittami, S. Doulbeau, M. Elias, G. Farnham, C. M. M. Gachon, B. Gschloessl, S. Heesch, K. Jabbari, C. Jubin, H. Kawai, K. Kimura, B. Kloareg, F. C. Küpper, D. Lang, A. Le Bail, C. Leblanc, P. Lerouge, M. Lohr, P. J. Lopez, C. Martens, F. Maumus, G. Michel, D. Miranda-Saavedra, J. Morales, H. Moreau,

T. Motomura, C. Nagasato, C. A. Napoli, D. R. Nelson, P. Nyvall-Collén, A. F. Peters, C. Pommier, P. Potin, J. Poulain, H. Quesneville, B. Read, S. A. Rensing, A. Ritter, S. Rousvoal, M. Samanta, G. Samson, D. C. Schroeder, B. Ségurens, M. Strittmatter, T. Tonon, J. W. Tregear, K. Valentin, P. von Dassow, T. Yamagishi, Y. Van de Peer, P. Wincker, The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. *Nature* **465**, 617–621 (2010).

66. A. Cormier, K. Avia, L. Sterck, T. Derrien, V. Wucher, G. Andres, M. Monsoor, O. Godfroy, A. Lipinska, M.-M. Perrineau, Y. Van De Peer, C. Hitte, E. Corre, S. M. Coelho, J. M. Cock, 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).

67. A. F. Peters, D. Marie, D. Scornet, B. Kloareg, J. Mark Cock, PROPOSAL OF ECTOCARPUS SILICULOSUS (ECTOCARPALES, PHAEOPHYCEAE) AS A MODEL ORGANISM FOR BROWN ALGAL GENETICS AND GENOMICS1,2. *Journal of Phycology* **40**, 1079–1088 (2004).

68. A. Montecinos, M. Valero, M.-L. Guillemin, A. Peters, A. Desrut, L. Couceiro, Species delimitation and phylogeographic analyses in the Ectocarpus subgroup siliculosi (Ectocarpales, Phaeophyceae). *Journal of Phycology* **53**, 17–31 (2017).

69. R. Starr, J. Zeikus, UTEX—The Culture Collection of Algae at the University of Texas at Austin. *Journal of Phycology* **29**, 1–106 (2004).

70. S. M. Coelho, D. Scornet, S. Rousvoal, N. T. Peters, L. Dartevelle, A. F. Peters, J. M. Cock, How to cultivate *Ectocarpus*. *Cold Spring Harb Protoc* **2012**, 258–261 (2012).

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

72. S. M. Coelho, D. Scornet, S. Rousvoal, N. Peters, L. Dartevelle, A. F. Peters, J. M. Cock, Genetic crosses between *Ectocarpus* strains. *Cold Spring Harb Protoc* **2012**, 262–265 (2012).

73. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676–682 (2012).

74. D. Ershov, M.-S. Phan, J. W. Pylvänäinen, S. U. Rigaud, L. Le Blanc, A. Charles-Orszag, J. R. W. Conway, R. F. Laine, N. H. Roy, D. Bonazzi, G. Duménil, G. Jacquemet, J.-Y. Tinevez, TrackMate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines. *Nat Methods* **19**, 829–832 (2022).

75. M. Berger in *Geometry, Band 1* (Springer-Verlag, 1987), p. 427.

76. F. B. Haas, N. Fernandez-Pozo, R. Meyberg, P.-F. Perroud, M. Göttig, N. Stingl, D. Saint-Marcoux, J. A. Langdale, S. A. Rensing, Single Nucleotide Polymorphism Charting of P. patens Reveals Accumulation of Somatic Mutations During in vitro Culture on the Scale of Natural Variation by Selfing. *Front. Plant Sci.* **11**, 813 (2020).

77. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

78. T. D. Wu, S. Nacu, Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881 (2010).

79. L. Sterck, K. Billiau, T. Abeel, P. Rouzé, Y. Van De Peer, ORCAE: online resource for community annotation of eukaryotes. *Nat Methods* **9**, 1041–1041 (2012).

80. M. Haeussler, K. Schönig, H. Eckert, A. Eschstruth, J. Mianné, J.-B. Renaud, S. Schneider-Maunoury, A. Shkumatava, L. Teboul, J. Kent, J.-S. Joly, J.-P. Concordet, Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* **17**, 148 (2016).

81. C. Diesh, G. J. Stevens, P. Xie, T. De Jesus Martinez, E. A. Hershberg, A. Leung, E. Guo, S. Dider, J. Zhang, C. Bridge, G. Hogue, A. Duncan, M. Morgan, T. Flores, B. N. Bimber, R. Haw, S. Cain, R. M. Buels, L. D. Stein, I. H. Holmes, JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biol* **24**, 74 (2023).

82. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525–527 (2016).

83. C. Soneson, M. I. Love, M. D. Robinson, Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2015).

84. K. A. Bogaert, T. Beeckman, O. De Clerck, Two-step cell polarization in algal zygotes. *Nat Plants* **3**, 16221 (2017).

85. T. Yamagishi, D. G. Müller, H. Kawai, Comparative transcriptome analysis of *Discosporangium mesarthrocarpum* (Phaeophyceae), *Schizocladia ischiensis* (Schizocladiophyceae), and *Phaeothamnion confervicola* (Phaeothamniophyceae), with special reference to cell wall-related genes. *J. Phycol.* **50**, 543–551 (2014).

86. W. J. Hatchett, A. O. Jueterbock, M. Kopp, J. A. Coyer, S. M. Coelho, G. Hoarau, A. P. Lipinska, Evolutionary dynamics of sex-biased gene expression in a young XY system: insights from the brown alga genus *Fucus*. *New Phytologist* **238**, 422–437 (2023).

87. K. Katoh, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**, 3059–3066 (2002).

88. M. Remmert, A. Biegert, A. Hauser, J. Söding, HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* **9**, 173–175 (2012).

89. M. Steinegger, M. Meier, M. Mirdita, H. Vöhringer, S. J. Haunsberger, J. Söding, HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 473 (2019).

90. M. Van Kempen, S. S. Kim, C. Tumescheit, M. Mirdita, J. Lee, C. L. M. Gilchrist, J. Söding, M. Steinegger, Fast and accurate protein structure search with Foldseek. *Nat Biotechnol*, doi: 10.1038/s41587-023-01773-0 (2023).

91. D. M. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**, 157 (2015).

92. T. Abeel, T. Van Parys, Y. Saeys, J. Galagan, Y. Van de Peer, GenomeView: a next-generation genome browser. *Nucleic Acids Research* **40**, e12–e12 (2012).

93. R. Zaru, S. Orchard, The UniProt Consortium, UniProt Tools: BLAST, Align, Peptide Search, and ID Mapping. *Current Protocols* **3**, e697 (2023).

94. E. Quevillon, V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, R. Lopez, InterProScan: protein domains identifier. *Nucleic Acids Res* **33**, W116-120 (2005).

95. M. Steinegger, J. Söding, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35**, 1026–1028 (2017).

96. D. J. Richter, C. Berney, J. F. H. Strassert, Y.-P. Poh, E. K. Herman, S. A. Muñoz-Gómez, J. G. Wideman, F. Burki, C. De Vargas, EukProt: A database of genome-scale predicted proteins across the diversity of eukaryotes. *Peer Community Journal* **2**, e56 (2022).

97. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

98. V. Thumuluri, J. J. Almagro Armenteros, A. R. Johansen, H. Nielsen, O. Winther, DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. *Nucleic Acids Research* **50**, W228–W234 (2022).

Acknowledgments

We thank Agnes Henschen for the help with DNA extractions, and Andrea Belkacemi and Dorothe Koch for assistance with algal cultures. Daniel Liesner shared HMG and APT sequences from *Laminaria pallida*.

Funding

Max Planck Society

European Research Council grant 864038 (SMC)
Moore Foundation (SMC)
Alexander von Humboldt Foundation (JBR)

Author contributions

Conceptualization: SMC.
Methodology: SMC, RL, MH, SC, YB, MR, CM, VA.
Investigation: MR, RL, CG, MH, SC, APL, FH, JBR, VA, MZ, AH, CM.
Visualization: SMC, VA, JBR.
Funding acquisition: SMC.
Project conceptualization and administration: SMC.
Supervision: SMC, VA, JBR.
Writing (original draft): SMC, RL, JBR, VA.
Writing (review & editing): SMC

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data are available at the NCBI BioProject PRJNA1022600 and in the main text or the supplementary materials; accession codes are listed in Table S3 and S4.

Supplementary Materials

Figs. S1 to S13
Tables S1 to S12
Movies S1 to S5
Supplemental Dataset 1

Supplementary Materials for

Repeated co-option of HMG-box genes for sex-determination in brown algae and animals

Remy Luthringer[1†], Morgane Raphalen[1†], Carla Guerra[1], Sebastien Colin[1], Claudia Martinho[1], Min Zheng[1], Masakazu Hoshino[1], Yacine Badis[2], Agnieszka P Lipinska[1], Fabian Haas[1], Josué Barrera-Redondo[1], Vikram Alva[1], Susana M. Coelho[1]

Corresponding author: susana.coelho@tuebingen.mpg.de

The PDF file includes:

Figs. S1 to S13

Tables S1 to S12

Other Supplementary Materials for this manuscript include the following:

Movies S1 to S5

Supplemental Dataset 1 (protein sequences used in this study)

Figure S1. Multiple sequence alignment of representative HMG-box domains from *Homo sapiens* (Hs), *Mus musculus* (Mm), *Phycomyces blakesleeanus* (Pb), *Neurospora crassa* (Nc), *Schizosaccharomyces pombe* (Sp), *Ectocarpus* sp.7 (Ec), *Dictyota dichotoma* (Dd), *Schizocladia ischiensis* (Si), and *Saccharomyces cerevisiae* (Sc). Residues conserved across all sequences are highlighted in orange, those conserved in SRY, SOX, and fungal MAT proteins in blue, and those conserved in HMG-sex and HMGB family proteins in green. Grey highlights indicate residues involved in DNA interactions in SRY. The secondary structure (h=α-helix) of the SRY HMG-box domain and the starting positions of all HMG-box domains are also indicated.

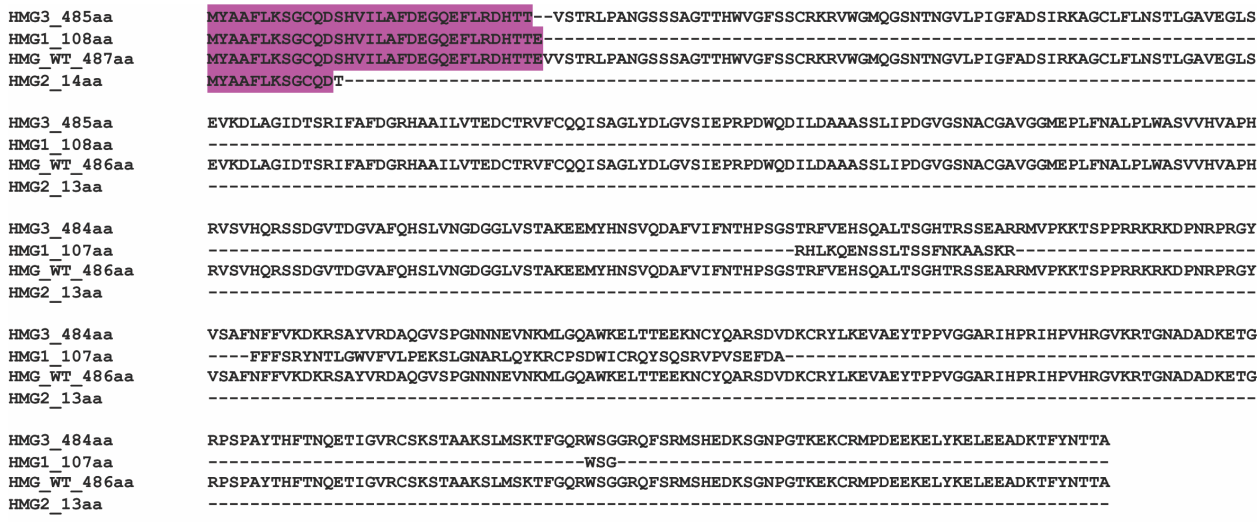

Fig. S2. Protein sequence alignments of *Ectocarpus* wild-type (WT) and mutants.

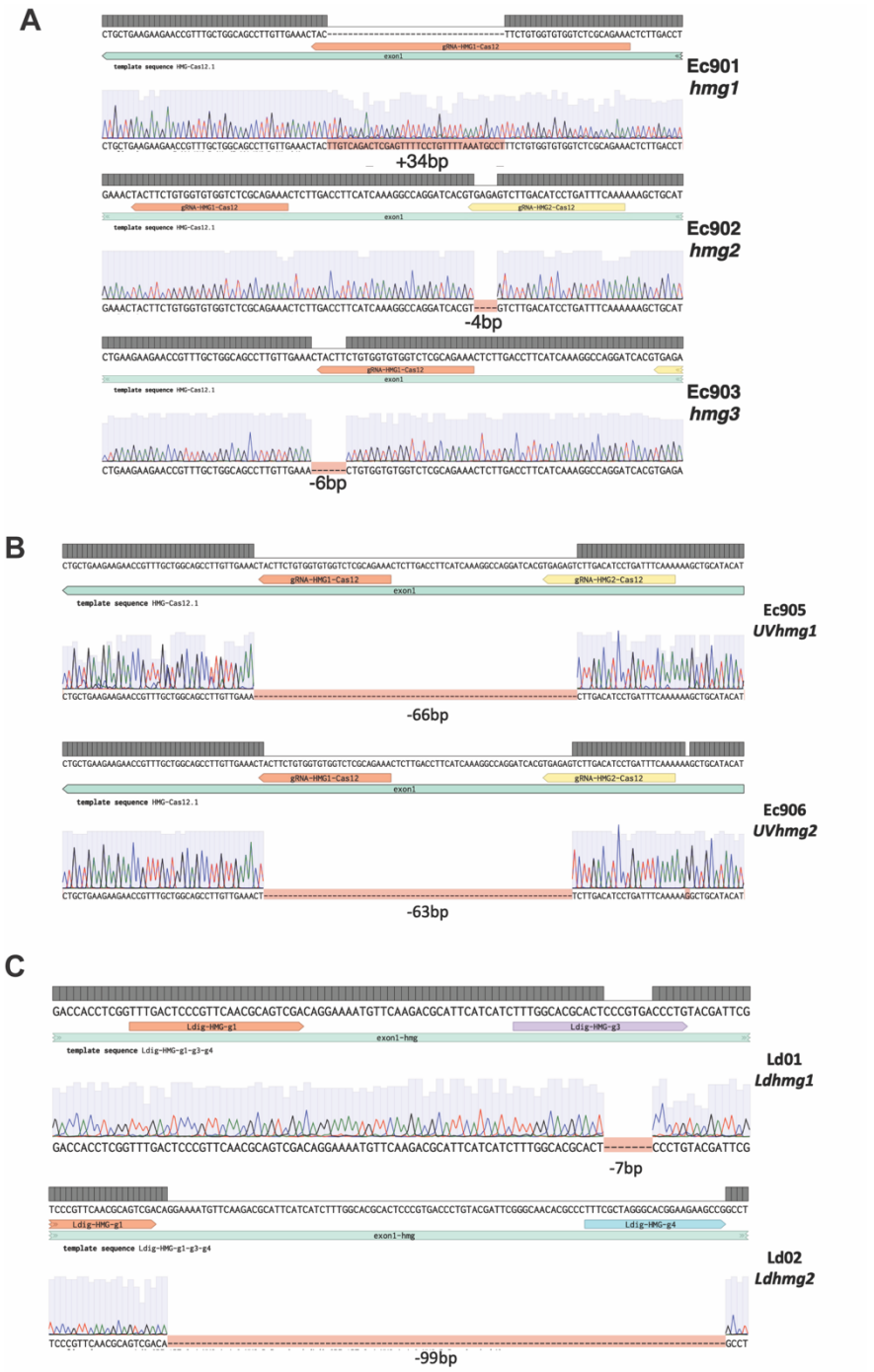

Fig. S3. (A) Sanger sequencing chromatogram illustrating genomic HMG deletions in Ec32 background. (B) Chromatogram illustrating genomic HMG deletions in the Ec581 (UV) background. (C) Chromatogram illustrating genomic HMG deletions in *L. digitata.*

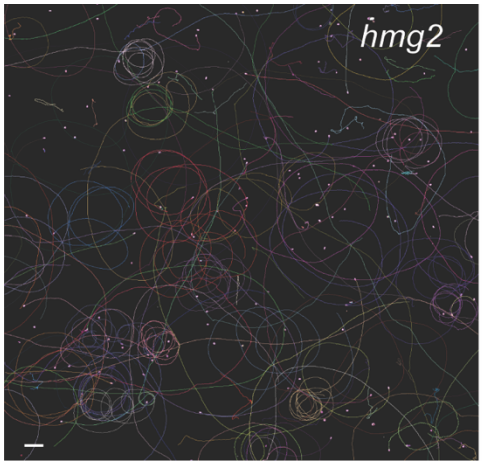

Fig. S4. The movement of zoids in absence of females is not affected by mutations at the HMG-sex locus. Trajectory over 15-sec of HMG KO (*hmg2*) zoids in absence of female gametes. Note the large circles and linear pattern of the zoids, similar to control WT gametes. Colors represent different zoid trajectories. Scale bar:10 μm.

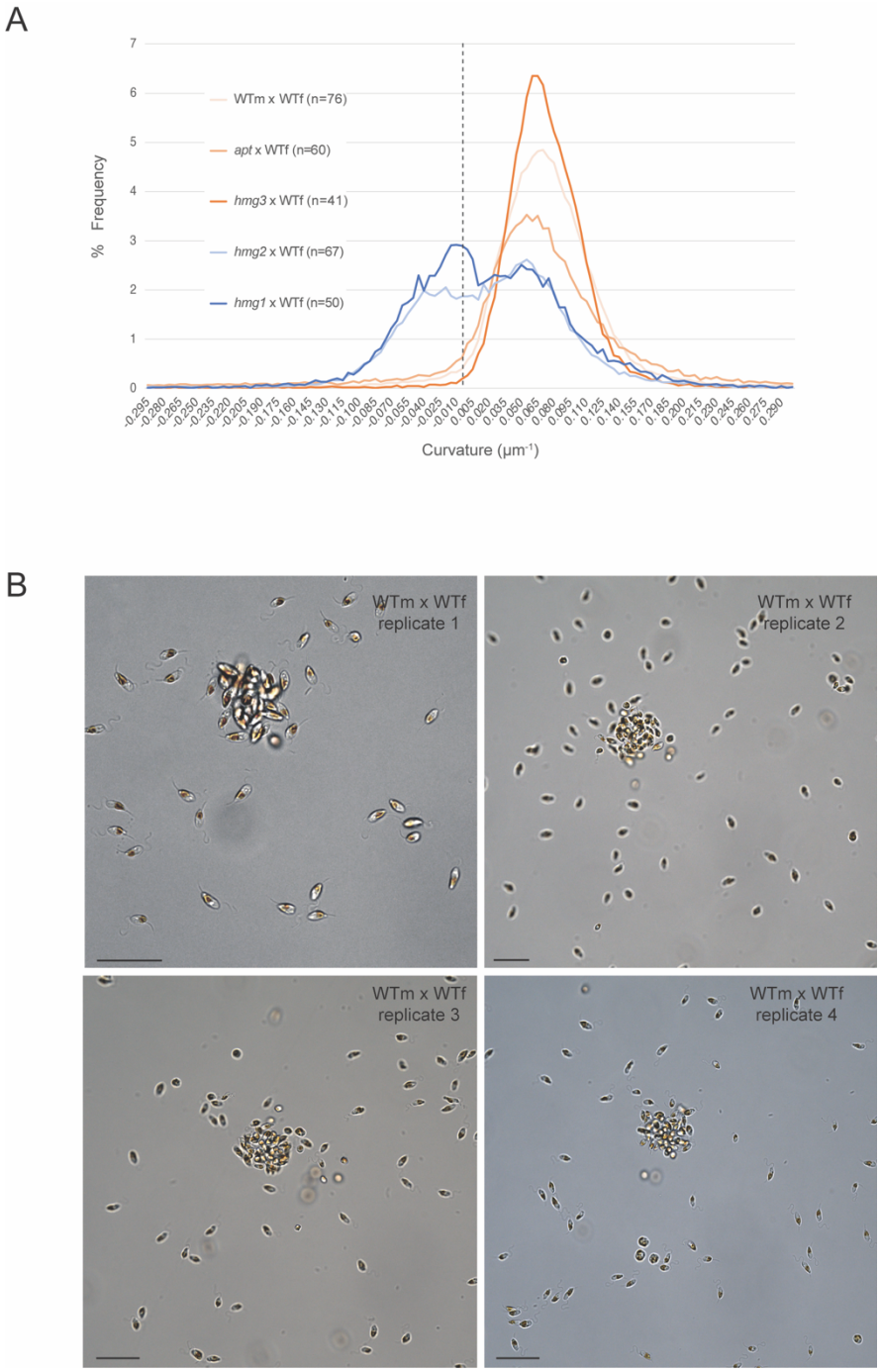

Fig. S5. **Zoid swimming trajectories in WT and mutant samples.** A) Quantification of the differences in the curvature of the swimming trajectories of WT and mutant zoids when confronted with WT female gametes. The plot represents Menger curvature value K (in μm^{-1}) along the motion trajectory of the zoids for each tracking experiment with the different mating combinations. *K*=0 (dotted line) if the 3 points are collinear, and K can discriminate counterclockwise curvature (negative) from clockwise curvature (positive) (Berger, 1987). The frequencies were normalized by the total number of timepoints measured for each distribution in order to make the distributions comparable. The number of examined trajectories (n) is indicated in the figure. See also **Table S12**. B) Male WT gametes are attracted to settled female WT gametes by pheromone production, forming typical 'clusters of attraction'. Examples of clusters of attraction (WTm x WTf). Strains used were Ec25 (female) and Ec32 (male). Bar:10 µm.

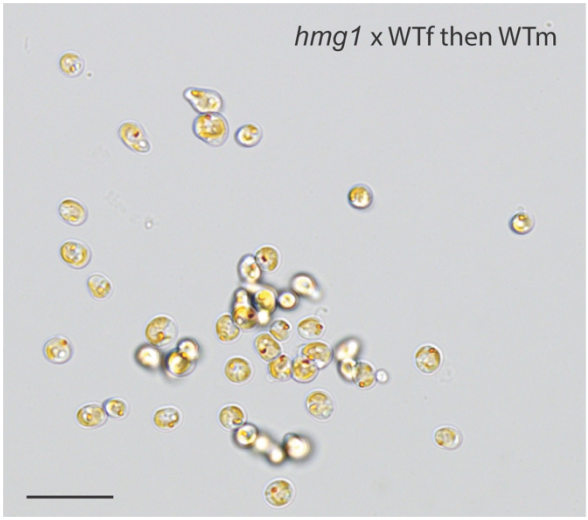

Fig. S6. **Female WT gametes are fully functional**. We tested if the female WT gametes were fully viable. Wild-type female (WTf) gametes attract WT male gametes and form clusters, but they fail to attract *hmg* KO mutant zoids. *hmg* KO zoids were not attracted to these same female gametes, indicating they are insensitive to the pheromone. Bar: 10 μm.

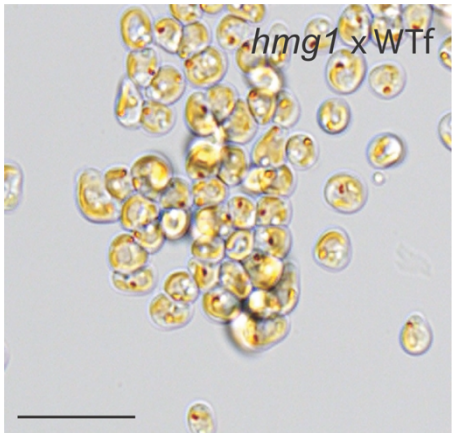

Fig. S7. **A functional HMG-sex is required for gamete fusion**. *hmg-sex* KO do not fuse with female WT gametes even at high gamete concentration that maximizes gamete encounter. Bar: 10 μm.

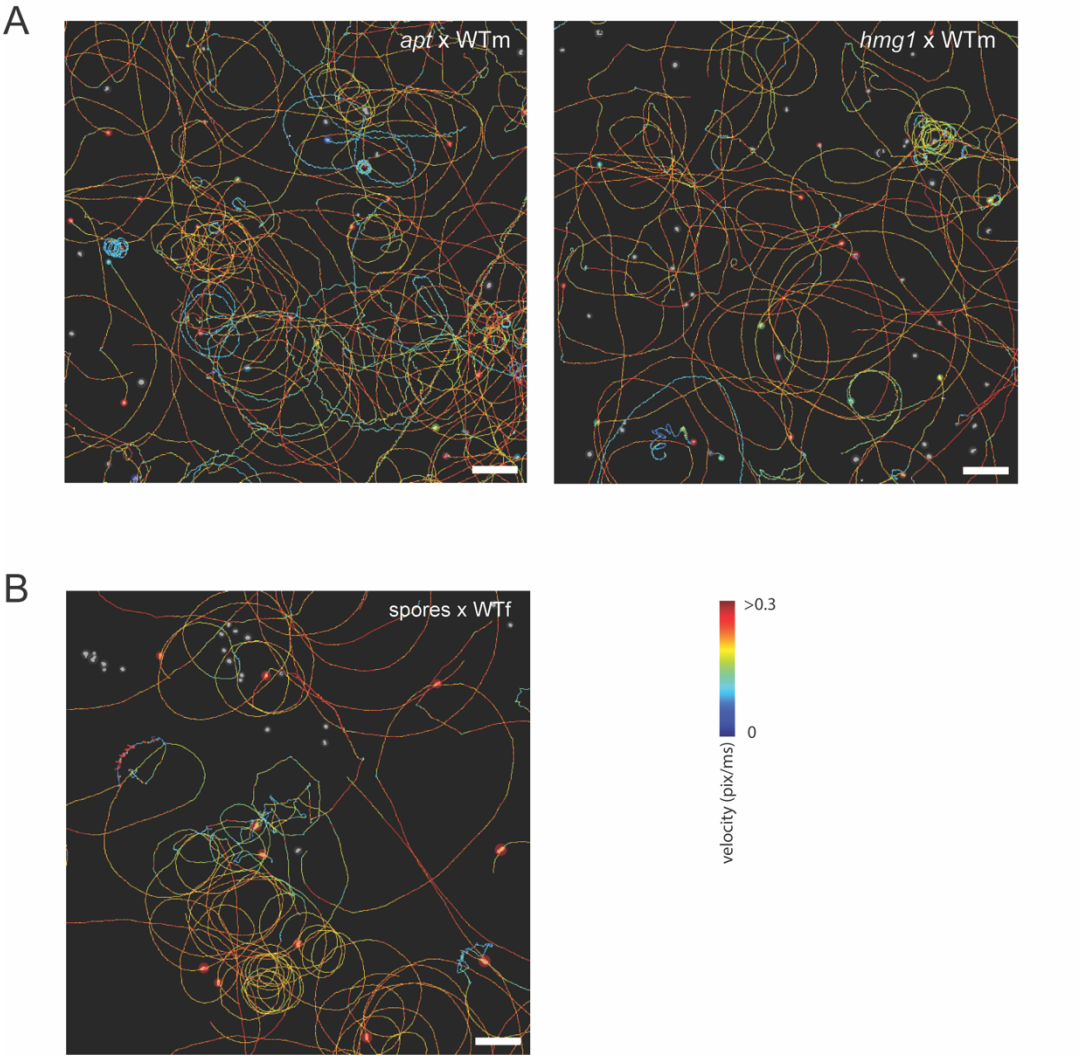

Fig. S8. Disruption of HMG-sex does not lead to sex-reversal. (A) Tracking the movement of gametes, colors indicate speed (pixels/sec). *hmg-sex* KO lose maleness but do not produce functional female gametes. Control crosses using *apt* and *hmg1* crossed with WT male (WTm) gametes. Mutant zoids swimming behavior is not modified by the presence of WT male gametes, and no zygotes are produced. **(B)** Wild-type asexual spores' swimming behavior in response to settled female gametes. Spores swim randomly in large circles, and do not respond to the pheromone. Scale bar = 50μm.

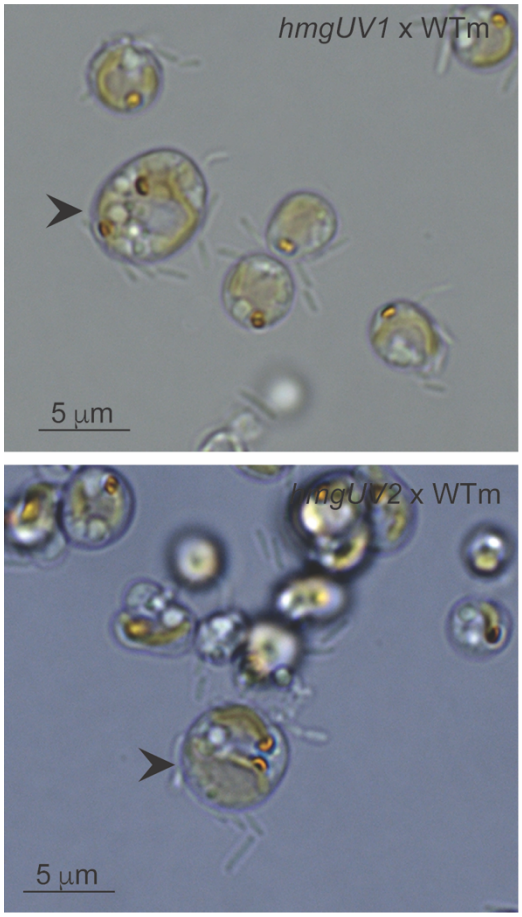

Fig. S9. Deletion of HMG-sex leads to sex-reversal in the presence of a U-sex chromosome. Test crosses to examine the phenotypic sex of *HMG-sex* mutants in a UV (2n) background. Note the large zygote (arrow head) with two eyespots (orange dots) formed in a cross between *hmg-sex* diploid UV individual against a male WT strain (Ec32). See also Table S10. Bar: 5 µm.

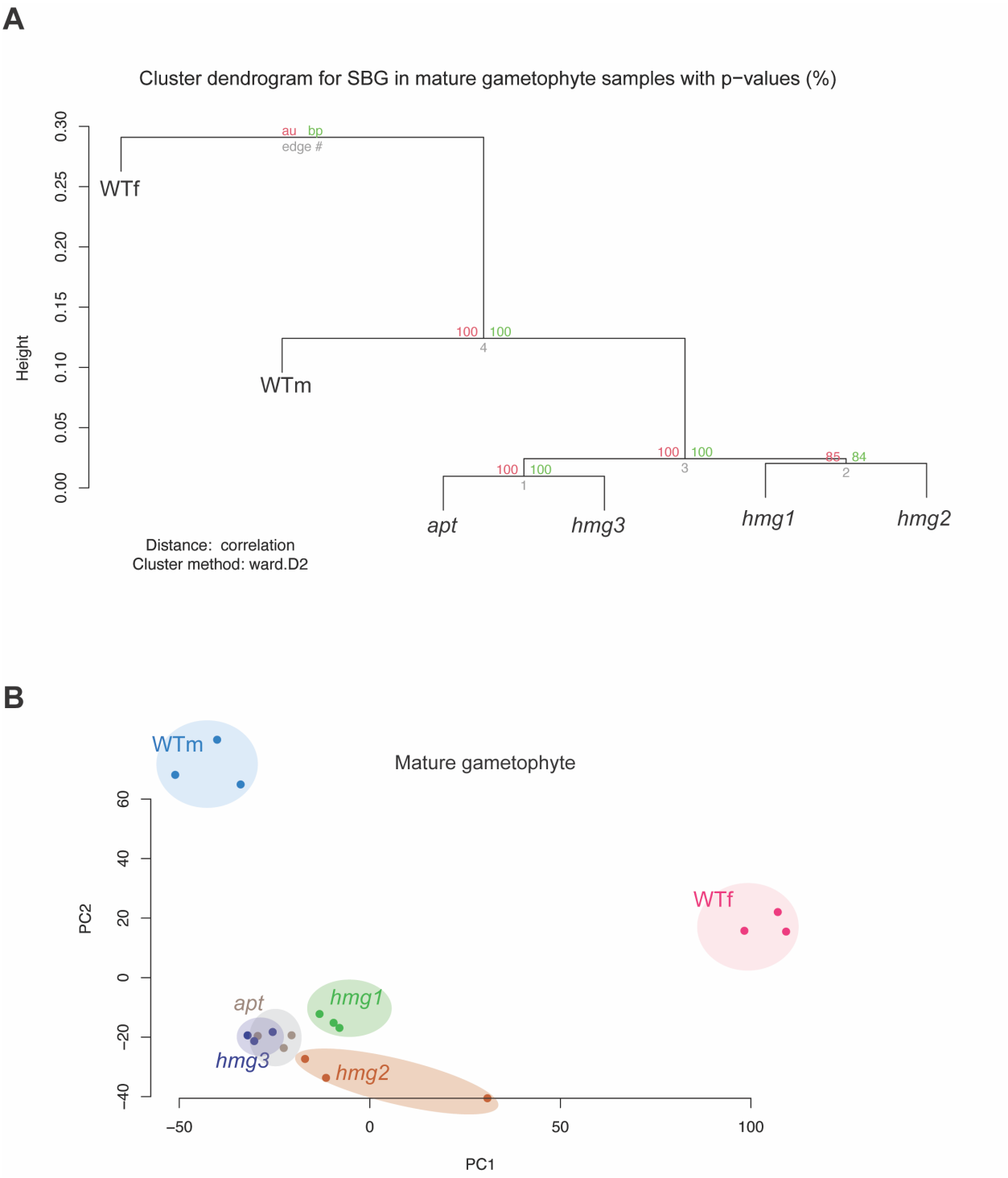

Fig. S10. Transcriptomic patterns of sex-biased genes in mature gametophytes of the wild-type and mutant samples. **A**) Cluster dendrogram based on all the log_2(TPM+1) expression values of the sex biased genes in the mature gametophytes. **B**) PCA (based on whole transcriptome).

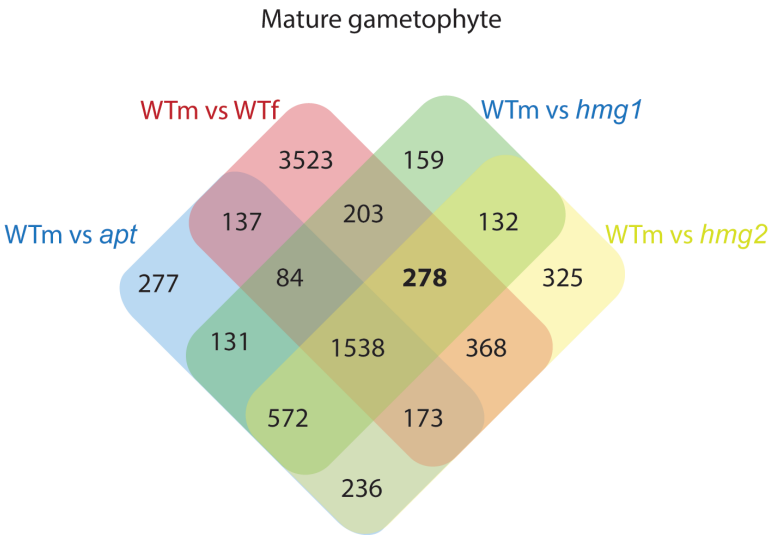

Fig. S11. **Venn diagram showing the overlap between the sets of sex-biased genes that have significantly different transcript abundances in WT versus mutant strains.** The set of 278 genes that were selected for further analysis correspond to the SBGs that are exclusively differentially expressed in *hmg1* and *hmg2* mutants (which constitute the putative effectors SBG, acting downstream of HMG).

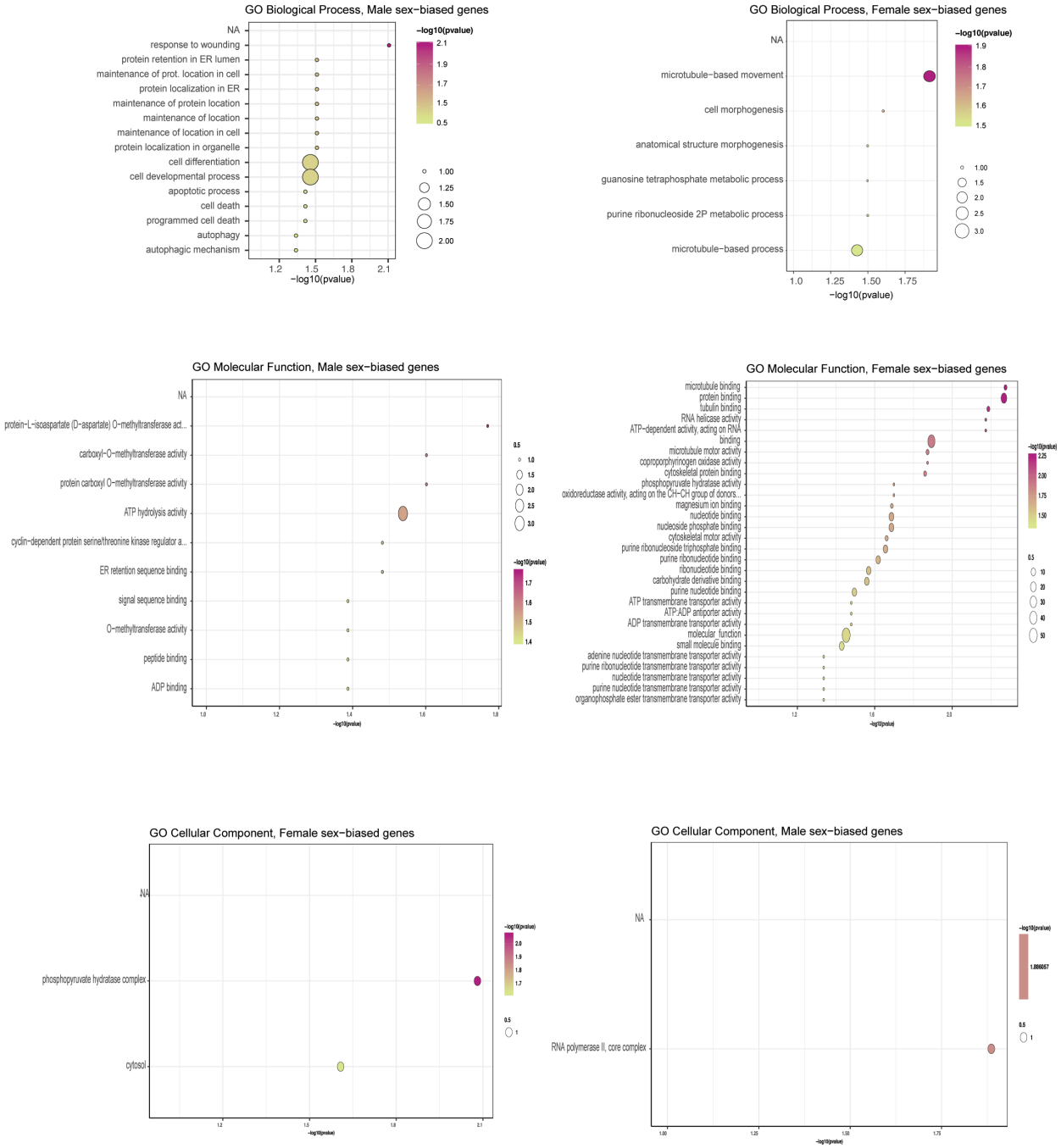

Fig. S12. GO term enrichment in the subset of male and female-biased genes that are potentially regulated by HMG-sex.

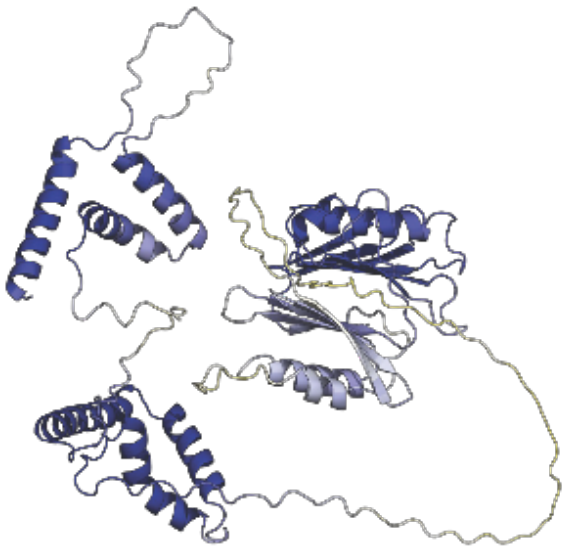

Figure S13. The AlphaFold2-predicted structure of *Ectocarpus* sp.7. HMG-sex is displayed using a yellow-white-blue color gradient to represent the pLDDT confidence scores. Here, yellow indicates low confidence, while blue signifies high confidence in the structural prediction.

Supplemental Tables Legends

Table S1. Strains used in this study.

Table S2. Samples used for transcriptomics and accession numbers.

Table S3. Predicted possible off-target sites for the crRNAs used in this study.

Table S4. Sequences of oligonucleotide primers and crRNAs used in this study

Table S5. Fusion success of male WT and mutant gametes with female gametes. 'Total number' corresponds to the total number of scored individuals (developing either by parthenogenesis or derived from fusion of gametes). Fertilisation success was assessed by counting the proportion of zygotes obtained after crossing either WT (Ec32m) or mutant (*hmg1-3*) males with wild type female (Ec25f). Statistical significance is represented by different letters above the plots (Wilcoxon test, $p<0.05$). Number of scored germlings is presented in brackets.

Table S6. Time of settlement (in minutes) of gametes and spores in the different samples. The time needed for 95% of the zoids to settle in the different replicate samples was scored. Male WT gametes of *Ectocarpus* swim for longer, whereas female WT gametes rapidly settle and start releasing the sex-pheromone. Each value on the table represents independent, replicated petri dishes.

Table S7. Statistical analysis for the comparisons of transcript abundances in wild type and mutant strains in the mature gametophyte stage of the life cycle.

Table S8. Sex-biased gene expression analysis using DESeq2.

Table S9. Gene expression changes in wild-type and mutant lines during mature gametophyte stage of the life cycle. Values represent $\log_2(TPM+1)$.

Table S10. Number of zygotes obtained when *hmgUV* mutant gametophyes (2n) were confronted with WT male (n) and WT female (n) test cross lines. A significantly higher number of zygotes were produced in the crosses between *UVhmg* mutant with a WT male compared to crosses involving a WT female line (chi-squared test, $p<1E-4$) indicating the deletion of HMG-sex in a UV (2n) background leads to the production of a female gametophyte.

Table S11. Sequences of oligonucleotide primers and crRNAs used for *L. digitata*.

Table S12. Quantification of the zoid swimming behavior. Menger curvature measurements along the trajectory of WT versus mutant zoids.

Movies Legends

Movie S1. Tracking the movement of gametes for 15sec; colors indicate speed (pixels/sec) as in Figure 2. Cross between *aptHMG* x WTf lines.

Movie S2. Tracking the movement of gametes for 15sec; colors indicate speed (pixels/sec) as in Figure 2. Cross between *WTm* x WTf lines.

Movie S3. Tracking the movement of gametes for 15sec; colors indicate speed (pixels/sec) as in Figure 2. Cross between *hmg1* x WTf lines.

Movie S4. Tracking the movement of gametes for 15sec; colors indicate speed (pixels/sec) as in Figure 2. Cross between *hmg2* x WTf lines.

Movie S5. Tracking the movement of gametes for 15sec; colors indicate speed (pixels/sec) as in Figure 2. Cross between *hmg3* x WTf lines.

Supplemental Dataset 1. Protein sequences used for the study.

Ectocarpus HMG-box domain-containing proteins

>Ec-15_003270.1 High mobility group domain protein (200) ;mRNA; f:3499495-3501540
MGLVGLTVHPAAGTVAAVDKDGGCQSPETISRTALQGEASKKSAAGEKRKTSSLSSSSTS
AAAAASPWSISKKLLEERARKAEYKAGRELVSKMSRKDSARPKGPQSSYIVFYTEKMASF
KAARPGMSITDIATAVGEAWRRLSDEMKLPYTKKAEADRKRYEDQLRAIARSAATSAATT
TTGAAPTATIAPTSFLRNV

>Ec-23_002100.1 High mobility group and SAP domain protein (289) ;mRNA; r:2172381-2181706
MATTATATVRSSRDKNAVERFVPPTAKAATDKNALYEGSGTKLADIPIVPVNLNKIDADH
DLLKAFHAFLYGGVGKKTVRKKNIRSFSGFPDDDGKDARMKKLTESKKWTVAALRDLCTL
LGLEKGGDKAEVVDRIVSFLASPDPSESKGPVAKKPKKAAKKRKSTGKGKKTKKAKREVK
PRPLNGYMMFTREARAKVLKADPGLKPTEVTAKMGQMWAECSDAQKEKYKQQGIKEFEKK
KGAMAEEEEDEEDDEGEDEGEDEGEDEAEDEGEGEGEDEDEDSGDDSE

>Ec-00_005550.1 High mobility group domain protein (353) ;mRNA; f:8849812-8851762
MWFSRHHRAILRTKNREMSFSDIGKAVGVLWKSATDEQKRPFDDLSAEDKARYARELAEY
KAWLLLKQDSTENKGEVPTQADKAAAATTAAAPTSRHASVAPPAPAIAAAAAAATTTTTT
TTPAAELSVPSTITTAATPVAPAPTPAPPMTEANVATSERGTATTPVAPGAALPALPPPP
LPPASLPTSFPEHPSHDGGMGVAMMHQQAKRQRRGESGAGGGDCSIPGAGAVPGVMAGVA
AGAAVEGIVAMTPMDQESLAARSRSHVPHLIQQQQQQQQQQQYQQTQHSIMQHQRGQQQH
QHQHHQYQQQQQQQQQQQQQQQQQQHQHQHQLCDTTRVEGPMVVSQPSLHYSV

>Ec-01_000710.1 High mobility group and PHD zinc finger domain protein (1572) ;mRNA; f:521316-529428
MDDEETQFLLDDEDEKPRKEKNQKKTPASPKSKGNETAAVVQAGTAGMSRAKGPKTGPGS
RGGKGGKGAKVGAGAKGDADKHDPAKPKKARSAYNFYLLKRIAQLKEEGVVEQHRDRFAQ
AAGEWRDMTFGERIPYEDMAKADKERHQKDLDIATRNLNAALHGTVPKHGQGMCSAIGFP
GLKALAANSPHEDVCAVCKEEGDLLCCDFCTSTYHLTCLDPPMLSLPSDDVQWACPACSA
SIEVAEMSAPQLKPKRERKDGRKRQRTSTAAVAGSSSPRAGADAGGNFTIPKLKKPGKKR
NKAAAAAAGEEGEAKLSPRIVKREEEKDELTTAAAVLASSANSANRNRMVKAPQKAEISP
RVGLGLTASTDEMSAATLRYPDLSGSRNNDKSPTKAMMSLAVAAASVAANEAAIAAAVAA
AAAAPAAAPPPPPKALGLKKLLHQRVQHDEANSNPAFSPKKLLKHRLLAAQQATTTTSSD
PAPSPSPPASSPSTGGGGNSYSRGSDDAGSADRYAGQSGERGASGAKWRRSGFAGAYAGD
SGRNAGPYGGREARGSDERDRIAAEGELNGRVLADGAGYGRRNGTGGMNDNVGGRSENNR
GRHADDGTAEGSGVGGGDRAALDGERRRRAELGGGGYDAGRNGDANANGVHGREASRSGW
GAGSSGERNGRFAKGMNNPTDKWTGGGESGRMDAPRHHGFSPAGFEEEGGRRQGRQPQYR
GDSYSSGQGDATQQRRDDGGGERRDNSSSGGRGYNNKASALAPPPPPPPPHCWDGGGGGR
SSWRPDRPSEGKREPDGRTQLRESAGRGGFTDESRERQRRPRSREGEYANGVPQPGSCSR
SVDDERDGRGRPPWGGNALRGGSRSRAADGGGDRERGFSAAGRAGGQHYGSSNGRVGDSA
AAVAGGGGRGEGGSYGGAWSDRKRSLERFEGQDWSHGSPKKHAPSPRTGGAFRAESRGRP
ARWDDVRPNSGGGGSGGGRPDNREPSRDSRAPGEGGSREGRGSDARVCGTEASVGRARGG
GGGGGEQTAGAWGGQGGERSDRRDSGGSSTRAPEHGGARDGPSSSRGGSEGAAWQDAAGG
RGGHHNHHRQGGGGGGGGVGARGRGAVRGHWKARLAGGRGGSDAQRGHHGEGDRAREPSV
AAQQSGGRDNGWPERAGGGSAGRNDGNGGDPAAGSRQAGRSEEPRVAGGRAGEKGREAGG

SVHAPGSRYDSQARGSSAGLGGIGADRRTKGPTDFRYGQTAAANGGTGGGGRGAGSRKEK
DYSGFALPKPGAPPHASSSAGPPASGGGTGYAAFAAGSSSSPSSSSAATTTTSASVSRPD
RSARSQEFGQDNQQQHQQQQQQQNRSMARPNGGGRGGGNGGGSQSSFHPRGMSERGGGRG
FRQGGGGGGVAKDFQASRGDGVNDRGNYSRSSARSLSPPAASRSARRSSASPNGRQQGYK
RQGNSGPDACSPSGGGGDDAYQRSGKPRPGWESSGMGRGGAGGRRPWKGRGEYHHHSNGG
KEEERYGNSGHASNNSNGRFQNSDHRSRGDFPGRGEGGGFGGRGRGWGGGRGWGGGRGRF
DSRSDNGSNS

>Ec-03_001170.1 DPB3 class protein with histone-like transcription factor (NF-YC subunit) and high mobility group domains (394) ;mRNA; f:1424978-1430031
MAGPPLIPSSSSNSNANVKKDDKPKDKPCPKRRRKSIDDTTTPTAPRTSTAAAPNKDGAG
PKVTQEKEQDHGETASADADAVISRTPLPEDSSAAATKEGQEGQGGGDDADGMAVDATET
AGRTTDAEQAEGAEAEQGEEGQDHEEEEEEEEEEEEEEEGEGEAGVAASPPSGEKPDGSK
IARARSAWMLFLADNREGVRKEHPGLAVGPMQKILSEMWKALGEAEVARYAKLAAEDKER
VKNELAAAGLTKLPTKNSSSAAVAGGPTSLVLPLARVRKTVRLDPSVGNISKEGLLLVTK
ATEVFMAVLADHAWNIGRQTGRKSVRPCDVADFVFAAPEMYWLKDEFREERSSVKKKQSR
PEAGGAKAAGATKAGVTAVPKGAKPITSFFQKA

>Ec-03_001500.1 GREB1 and High mobility group domain protein (2249) ;mRNA; f:1727898-1757376
MRLFPCPLRTGKQPPLSVGKALGLWHPTPFVEIENYLDLAVSFSLAHLLCSSMPHQRKVR
RTTAPPPEPCRRVRGATIAIFSWMPQCIGEHKQPGCEQDYRCNKPSMSSSRFCVDHQEQD
IEVFEEMWRNPATPDSGQGGGGSRYSGGNSSGGGGIGSVGGGGGGSGYSGGGGGGRRHQD
REAVNGARPDIDDSGDDEPQGAIQRVALIFSASDTGTVGEEFVGGLIPAVRVVEELDVGP
DVPVASKRKRLVDRRERLVHLCTAFRKLMPEASTCSEYSDVVDTLNAAFAGSAKGHQIMK
KAAESAAKNAGCEPHVLYDIRFPEGVIRVNYGEVVSARQQVDARSGNTLSTVLIETDDET
KSVLVVVLGGPLANLLKTGQENLNVDGYYLDEGPRQASVFKVTVDISLKSGEGDDIVYSA
TIMADVHVVRGTLLAEFCEAQKWRGQLDLNHYLQEPILDESTLQVVCNGMACNEMVQVVQ
DIEDTTPVYTQCRISDDQDQHVAQLLKVRLNITAAEFGHLTAALQETGSPDSHNIQHVCW
PKQQEVAQRLFDKTNEVLGSRCISEFKGVEFMVLTPAFQTLQQQTINRLAKCGVLPQGKV
ASECVIMYIKSAKHDQLFERFKSRVRDNKDILFVLVADECHYGCTDGGAHDKYVNDEDLH
RHANLVLLGVSATPYNCLTDKSRVHESNVIEWFEENHPREDGIYRSMEFYLGTIPFRPDP
SACTLHVQVVPKTGQGSSFRVDIKDKAFTAPNEVAEALSAEIAVKLRECDANASMEISYC
KRGKRFLVKNKSRLFTVRLETNEASILAKLGFTTNLPLELGPDSEIRADSDWEIGNHDQP
GGRSLEDQRIRDDPGFTELRARFNRQFGPRGGKQRRGKNISTPTENDRQPSVALHDGHLL
LAEYLFSLVCLSMFRQEGLTMRMRSPGIGSETDNFLDELVTEDCLDSFDTRLGRVCRFNV
GTKGDPCFCDIHSVVDFVKQAYVSAAREELKDNPHDDRSCLRYILREKLKASAVAENTSA
EYSWFNETDRVLRDLLRGDSDRGHGRMVVMRVYEREEQLSMQHCLRQARRELLLHDGPTE
RPLFSVVCDNGDTNLSQDIEEYFRDSFNVQDDAGGVSTLGNRMVELRKLPGSRNTSLSYE
DLLNVPMILILCEKGRMGDTFPHSLGCFDLRLRVAGYFSSFEQELGRLCRYPAFRKINGG
SGPSREDAMSLGCEVLDRKRLVKVSTPAGILSVADTWWLLESIIASIGADDVITLEESVF
PLPTALVPSHTMGKLLKAVEEHRLTNTPISKCIRMAKMDLHMIGSTNIKKKPNTETENAL
HNYSAYRPAEKHFDHRRGLSERKDPRRLILSAECQIGKTGVYLDYLSKLTNATTASCSEI
VPCPPPVEDGWPREELSWLLPHWRTLIGIAPMSNEYSTLLASKYTEGIVRERIHLVRQSC
CEGNRWRDEYAALLVDSCGEHVRSVIGNQYIEALRQRELVPPFNTEGDPTPAPACFESLK
EAINWDGRFEPGLGVQFCACESQCTCAMAGDAYGKGLTLQLADLTRAGGEHGRVDERWNM
PQDGDEGPAAKDAHVRKFAAKQPPCNWANGEAQGNLSTCTFRHPREPESGHTFSLHVSIP
TEEIGVIFFNPKKTASRWIFTPSYNRAASGPRQAVLDRSVALQQGCDPPDVGTRFHINVL

VVRWEQFALYRRHFGHQYVVLAMPGRMPCFESEGGYCSAEEGGIGYARHFIQQWSSTRDI
PFVWMLDDNVQMCHELDVDTGEYNPCGFTHVMDSLERILMLPDSTNIVVNTAQQSVSKHA
ASRTTCPPDLESLRGQTVPGRTMHLTTVAGFCGKPDHYGVLGISRHGLGGENTKPTTTPF
GVNHSVYSFCLLNVRSTVKQRAYYPMKRYWEDVEFNHIVDEKGLVVCMFRKFSHSKKNLQ
PPHLQGPLLCPPFAFEHGDIGLDVLEETYEKVRRGSTMVIPTDYLNSLLKYLSCHVLGSI
YAEAIAYPSDRRRVILVDTPFEVLPTVPEQGSPRIELQTVPNSQVFLIMLLGGVKNQSLG
LLKHVLTDKLVKQQLKEVVVVTRLPQRDGSTGNHIGEWLQEFNDPESVVEVGSKIVKPKG
GRLGHGLLVLSFSSTSEHRREDENKLMAIRDSPPDEDEDTDTDEPPPPGSNEVQSQPPPR
KRAKKDAPRAAMSAVVHFSQAKFAETKAENLDMEDTEIGIVLGEKWRGVDENIKAQFQRI
ADDDLKRYEKELIDYMAINKAANARDKG

>Ec-11_002940.1 DPB3 class protein with histone-like transcription factor (NF-YC subunit) and high mobility group domains (421) ;mRNA; r:3119495-3125011
MADTGAAGAPASSLVAAAATAAASAAAASAFVGADGGGDGGSSGNGASSPAPAAPQAAAA
ASSSSSQLPPMSHSVSNAAAAAAAAAAAATTTAPPTDATVQSPTTTVQQAGPLSPGLDGA
DPAPGGGGSGVGVEGVTGGGAGAAVSATAVGPEETLAAAVPAAATPAAVPQAAVPQAAVP
QAAVPQAAEPTAGPGAVAAAPTPATTAAGPAGAGVGPISAAGPNGVAPATSTTGGGAGTS
SGKKRAASGGQSKPQSAWMLFIQERREQVKRDNPDIAVTQQQKIMSEIWKTLASDEREAY
RTRARDDQKRYKAEQADKPVVKKEEFYTNDEKGQSTELVCPMGRIGRIIKLDQDVQSVTK
EATALIGKAMELFTSLMAKESFSVAQSQNRRLLKLQDITDIMHNQDQFYWIKDDFPLPTR

>Ec-11_005660.1 High mobility group domain protein (243) ;mRNA; r:5625070-5628663
MQYEGEEHVAYDGDGGLDQGHYDDGSEGGAAGGAAPGSSKRRKKRAYKKAPDAPRRGRSA
YVLFSMEAREEVKNALPEGSKVTEVMKGIAAKWRELSETDKEEWTAKAAQDKDRYEQELS
VYDGPLKVPNKRAKKDPLAPKRAMSAFLHFSQSMRPRLRETYPEAKNMDMSKMLGQEWNR
MSDEEKLPYQTKAHDDTLRYREAMTVWKDGGADALASHMAARGETDGAVGKYEEDGDGGD
RT

>Ec-12_007830.1 High mobility group domain protein (125) ;mRNA; r:6925324-6928480
MAKPAVSKKSAKAASNGGPVKRKRKAKKDPNKPKGASSAFMQFSQKERAVVKQENPDMKV
TEISKVLGARWREMDDNDKAPFQKKADKDKARYQKEMAAYNAKKAAEPASEEEEEEEESD
EESD

>Ec-12_008000.1 High mobility group domain protein (646) ;mRNA; r:7086414-7095497; FACT complex SSRP1
MNFSKIGVQGYSLGNLQLTAEELNWTSDDKSSSKKVKWLDVSHATWAQYAAYCHLRLFMK
RDARPVRLDGFSKAQHADIEKFLSERDVSLSNESPNPGGGNYGDIEVLDNIVRFSSGNKT
LFDLNIKDVSQCVMPGVKKSNDVELQLHESDATDQTEDTLVAIRVTLPEKDEDTDERSPA
ESLQMAVMERANIHDVKGKVLVEFNESQGTFDFPRGRYSVEMYSHFMRMHGSRYDYKIQY
NDISKLFLLEKPDERYVAFVISLDKPIRQGQQKYQHLVLRTTKDEATITVNMSAEDLQKK
YDSNLNSEMIGPLHNLIAKIFKVLSNKPVYVTGKFSSTNGAKAVKCALGANEGYLYPLNK
SFIFIHKPTCIIGFDEIESVEFQRYGGAQGAGVTRNFDLCVAPKSVAGETPKPYVFSGID
RSEYSSLYSFLSTKKLRIKNIKESGNDNAMLQLGDLDDHDPYKAALDDDQGEDEESEDDA
DYAPDARGGSDAESSSDSDGDEDSDDRAHKTSKKRPGTGSAKKGPTQKKRASGGGSSKPA
PKKKRAKKDKDAPKGAMSAFMQFSQANRAQVKTDNPELKVTEISKVLGEKWGKLDETQKK
PYQDKADEDKARYKRERDAYDSKKAATEPPQQSDSNDSDAASDSD

>Ec-13_001750.1 High mobility group domain protein (487) ;mRNA; r:2861582-2913848 HMG-sex

MYAAFLKSGCQDSHVILAFDEGQEFLRDHTTEVVSTRLPANGSSSAGTTHWVGFSSCRKR
VWGMQGSNTNGVLPIGFADSIRKAGCLFLNSTLGAVEGLSEVKDLAGIDTSRIFAFDGRH
AAILVTEDCTRVFCQQISAGLYDLGVSIEPRPDWQDILDAAASSLIPDGVGSNACGAVGG
MEPLFNALPLWASVVHVAPHRVSVHQRSSDGVTDGVAFQHSLVNGDGGLVSTAKEEMYHN
SVQDAFVIFNTHPSGSTRFVEHSQALTSGHTRSSEARRMVPKKTSPPRRKRKDPNRPRGY
VSAFNFFVKDKRSAYVRDAQGVSPGNNNEVNKMLGQAWKELTTEEKNCYQARSDVDKCRY
LKEVAEYTPPVGGARIHPRIHPVHRGVKRTGNADADKETGRPSPAYTHFTNQETIGVRCS
KSTAAKSLMSKTFGQRWSGGRQFSRMSHEDKSGNPGTKEKCRMPDEEKELYKELEEADKT
FYNTTA

>Ec-12_007350.1 High mobility group, SNF2 and SLIDE domain protein (1467) ;mRNA; f:6538774-6568628
MSEIKAAASAYHLYMKEKHAEVKASLEAKGQTADFGDVMRELSSRWKELGSDDADRVKFE
TLAAEDRARFERESAAKDLEVAEAQRAKRAERESLTTESRMRGRPKSEQPKPKAKAMGPP
RQLSEEEKARRQERQDLKNKEKAERNARTQIAEAQHTSIKDEIAKQASARLQFLLKQSDI
FQHFGVGKQATQAGEAKQEKGAASAEAKAGEVTSPGKRRREAGKGGGAAAAAAAAEEEEE
EEEAPETTFLTKQPDCIKFGKMRHYQLEGLNWMIRLNDNGINGILADEMGLGKTLQSISV
LAYMHEYKGISGPHIILVPKSTLSNWLNELKRWCPALRPLRFHGTREERASLIEERLRVG
HNDRDWDVGGANLEKRSLQNIAWRYLIIDEAHRLKNEASMFSQTVRSFNMQHRLLLTGTP
LQNNLHELWALLNFLLPDVFSSSEQFDQWFNLEIDDKEAKENIIHQLHKILRPFMLRRLK
ADVEKSLPPKTETILYVGLSSKQKEVYRNVLLRDIDMVNGTGGGGNAGRTVILNIVMQLR
KCCNHPYLFAGVEDRKLDPLGDHLIINCGKMVLLDKLLKKLFDKGHRVLIFTQMTKMLDI
FEDFCVMRRYEYCRIDGNTSYESREDCIDAYNKPDSTKFVFMLSTRAGGLGINLQTADTV
ILYDSDWNPQADLQAMDRAHRIGQKRPVSVYRLVTENTVEEKVVERAQQKLKLDAMIVQQ
GRLTDNAKKLGKEQLLDALRFGADKVFRSKDTSITDADIDAIMAHGKEKTKALMESKLQV
SDKGDLLDFSLDGGIATQVYEGVDYSNKANRADAGGGAVPFQFIDTGKRERKKIASYKED
TFYRQQQAAAIKRRTMMPKHLRLPRMDECQFYDKKRLEQLHRKEEQLFVEAKEKGDLPND
LTNYEVLPEPLLSEKVRLLSEGFADWSKAQYMIFVRASAKHGRGAYDRIAAEVGKAEKDV
ERYAKAFWEKGSTAFAPADWERHVKNVEKGERKIEEINRLMGATRSFISKFANPWEQLTF
QYTGTQGKVFNAHEDRYLLCLAHKYGYGNWDLVKAAIRRSDRFRFDYFLRSCSADALGKR
CEALMKAAEKEDAEWHKRQGTERGAGPGGSERLGEEERRRVAERERHNSYKDFQGKVEKE
TGKLTELHLDKSRLRESLLADGGASAGSGPAGGADSATGGGGKEKTGAGVSSAVDNVDRQ
LAYLVAQNQAGGSSSSRAQGQAKPVPDRLVGELARFACRSGSKGVDKIVTDFTAIHPEAS
KRQVEKKIQEVAYKEKRAGDVRDSWYIREGFESLLEEGLTDAEKKDDKDTNAKQPTKSEA
SGKEADSSSKVGSSTSSKKRRPHGDGERPTVTVRSPFYLFKHAERETAVQALRDRGVSDP
SSDEIKDKIKMMWRELSRDAQEEYFQKAEKTGWVEDQDSGKGGGSSSGGKRKASFSGDVT
STGSSAPGSPTSAFTIPKRKKVEPSA

>Ec-14_006770.1 High mobility group, SNF2 and SLIDE domain protein (1564) ;mRNA; r:6249181-6266861
MSGSGAGAFGAPVKAATSAYMFFQGDQYSNHKDEFAGLALGQQGAEISRRWRELPQEERS
RFDALAAEDRKRFQEESEARDAEVAARQEANRAARYADPASSTYMRERASVEPKKERKVT
REEDMSEERLEARRLAKEKRAKQKDARLAAEAESERQKSSIAAAAAATARKRIKYILRQS
DIFATHFGVALSDSDEEEEEKEKEKEEDASKGAAAAGGSPSKRRQAGRGAADEAMDDVNG
EDTAPTYLTKQPPSISGGTMRSYQLEGLNWMVNLQAQGTNGILADEMGLGKTLQSISILA
YMRDFQNVTGPHIILLPKSVLGNWQLEFKRFCPDIRVLRLSGTKDERAATIRNDLKPGSP
EDERDWDVLVTTYEVANIEKTYLNKIGWRYLIIDEAHRLKNESSLFSMTVRELTTQYRLL
LTGTPLQNNLHELWALLNFLLPTVFQDSEAFSKVFDLNVDDADKKQNMIKQLHKILRPFM
LRRLKKEVEKSLPPKEETILFTSMSEVQRKVYKGVLMRDIDTINGTSAGRTAILNIVMQL

RKCCNHPYLFPNTEDRNLDPMGEHLVENCGKMILLDKLLTRLKAAGHRVLVFSQMTRMMD
ILEDLMHMREYKYCRIDGNTPHDTRQDLIEEYNAPGSEKFIFLLSTRAGGLGINLQSADT
CILYDSDWNPQADLQAQDRCHRIGQTKTVKVYRLVTEDTIEEKVVERAQQKLKLDAMVVQ
RGMLQGEKKLEKDEMLAAIRFGADAVFRCKDTVMSDQDLDAVLERGAKKTKDMQSKLNVA
EKGDMLDFSFDDGTGVQNFEGINYGDRKLRDQMRAGFIDIGRRERKPATVKAYAPPPPRV
EPKKMVLKMPKELEMPKMKDHQMFDRARIVELQDELQDNFKALKDRGTKIPVDGADIASM
LLSADKVKERAELIAQGFPGWTRGHYYSFIDGMAVYGRDRLELVATMVGKPAEEVVRYGE
VFWRQGPKTFSEEAWRRIKTRVETKEKRLNELNRIMMVTNELMEAVDDPWRNMEVRINQR
QGGAGTAGMQLQREWTKDEDRYLLCLTHLCGWGNWRKVRACILASPRFTFDYYLRSLSET
ALGKHCEQLMKSSEKYLQDLQDRMKKEAKREMEDRGALEKDEAAAAREDEEFKLRIRAAD
NMLGTAERELAEAESNKRNLEMIAEAIKSGTVSLDNPNVRSLLESCASNGIGTTGLNTKN
AVKKSAEGVAGAGAGGRKSPGVGGKKSASPGGGGQKKAVASKAAVPGNTGNGGGGKKTSA
ASSGTPAAAANGDAKKSGGGGGKGKGNAVAGKKADTTSSDGKGKVDAGKGGVKSPKASSS
TSNAVDKSPANSSGTKAKRTPTKSVGKPGLEVPKTLFPELVRLIEKGELTSKDDMLQIFQ
KKHPAFTQVSINKAIGVLGEKASRGSKWKVLDAGKKYLSMPVYQGNVPEEVAVLSALSAK
RKSGGSESHSAKKRKSTTPSKANGESKSPITNFFPPKTSGAADANGGGGSSAASKKRKAP
APAAPKSPGPSPGFVAFCRANRTRVQAELPIGSSKADTADRLLQLWGELDDKVAERWTSG
ALP

Representative sequences for each group in cluster map (Fig. 5A)

HMG-sex

>Ec-13_001750.1 (Ectocarpus sp. 7)
MYAAFLKSGCQDSHVILAFDEGQEFLRDHTTEVVSTRLPANGSSSAGTTHWVGFSSCRKR
VWGMQGSNTNGVLPIGFADSIRKAGCLFLNSTLGAVEGLSEVKDLAGIDTSRIFAFDGRH
AAILVTEDCTRVFCQQISAGLYDLGVSIEPRPDWQDILDAAASSLIPDGVGSNACGAVGG
MEPLFNALPLWASVVHVAPHRVSVHQRSSDGVTDGVAFQHSLVNGDGGLVSTAKEEMYHN
SVQDAFVIFNTHPSGSTRFVEHSQALTSGHTRSSEARRMVPKKTSPPRRKRKDPNRPRGY
VSAFNFFVKDKRSAYVRDAQGVSPGNNNEVNKMLGQAWKELTTEEKNCYQARSDVDKCRY
LKEVAEYTPPVGGARIHPRIHPVHRGVKRTGNADADKETGRPSPAYTHFTNQETIGVRCS
KSTAAKSLMSKTFGQRWSGGRQFSRMSHEDKSGNPGTKEKCRMPDEEKELYKELEEADKT
FYNTTA

>dis_TRINITY_DN9420_c0_g1 (Fucus distichus)
MYGAFISSGHHDSQVVLALENRAIFKTSKRPDGISLIATSISCGDCDAKGNTTLWIGVSS
RKRAWVTESFSREGIDPGALVDAGLRFLDETCGALDGLSTAECCKRTEMSRTMAYDGTNV
AVSVIDEGIRTFSQEIPSGLFNLGLSRKSRPSWQDELMTTVSELLGTNDFQAPRFTTIES
CMPTSPSPWAAVLHVTPFRVVVQQKRSDGGTEGLSLHFFHNALDTTLPETNRPQIVEDGY
IIFGPEDSGAKHTSSEPMIDKSNATIKRHKRDPNHPKGYISAFNFFVKDRRPSYVQNGQN
AQGRSLQHNNEINKILGRAWKNATEEEKRIYEEKSTIDKRRYFEEMRAYRPPEGYARTAP
RINMPKGCDGITLDQDLAAGSASHACRRPWPAYTHYAHQERKGVLSSGSNQAKTQMSKAF
GRRWRGVPSEEAVLYDELEDADKERYEREANAQDSAASGPG

>spi_TRINITY_DN144730_c0_g1 (Fucus spiralis)
MYGAFISSGHHDSQVVLALENRAIFKRSKRPDGASVIATSISCGECNAEGNTTLWIGVSS
RKRAWVTESFSREGTDPGALVDAGLHFLDETCGALDGLSTAECCKRTEMSRTLAYDGLNA
AVAVIEEGIRTFSQEIPSGLFKLGLSRKSRSSWQDELMTTVSELLGTNDFQAPRFTTIES
CMPTSPSPWAAVLHVAPFRVVVQQKRSDGGREGLSLHFFHNALDTTLPETNRPQIVEDGY
IIFGPEDSGTKHTSSEPMIDKSNAIIKRHKRDPNHPRGYISAFNFFVKDRRPSYVQNGQN
AQGRSLQHNNEINKILGRAWKNATEEEKRIYEEKSTIDKRRYFEEMRAYRPPEGYARTAP
RINMPKGCDGIALDQDLAGGSASHACRRPWPAYTHYAHQERKGVLSSGSNQAKTQMSKAF
GRRWRGVPSEEAVLYDELEDADKERYEREANAQDSAASGPG

>Dictyota_dichotoma_SP_DNT_23462 (Dictyota dichotoma)
MKIHRIQHLQSIRGDKEDTGAIPHLRWRIARGRTSWIGVASSGRAWACEASTSTNILPVS

VFQSLRDAGIFFLNMTCGAQHGLETSGNFGDEAESIRIFAYDGWEAAVLVRNGRAQVCCQ
RMSSGTFSLRSPGVHATEDDSAFHLAARHLFFGMESCGDRGAVGHEISTVFDSTSSKSQN
ALVHITSARVVIQQNDADGSKEILSVHRQDDATSRVTDHCVIFPRQPSSSLRLVPEVRAT
PSENARPAKRPKRDPNRPRGYISAFNFFVQDRRPAYVQSHPDSQGHSLEHNNEINKTLGK
VWKQATKQEKRIYEEKATADKLRYLHEMSFYRPPQGYERIAPRINLPKGFSALAWDQDLP
ACAAPQLCRPQPAYTHFAHQERNGVLSSGSSLAKTLMSKTFGQRWQGMPVDETELYFELE
QADKARYEEEACAQESAASWIP

>Upin_00015662 (Undaria pinnatifida)
MYVAFFNSGCQDSQVVLALENSAGFVRSQGQNGASVFSTNLTNSGSGTACWVGLSSRRRA
WAMEGSFENDVPSVDLVDGVRNAGLLFLDLDCEAIASLSAADDFDGIDRSRVFAYDGVHA
AIMVTEGRTRVFCQQVSSGLYDLGVSINPRIEWRNALGAAASSLFVDSMEGAADGTLGGF
SSPSERFSLSAWSSVVHMAPNRVAVHQRCTGGATKGLSAHYFRTMVENDPMQHGRKNSVQ
DAFVIFGRHSHDAVRLGPHPRALVHGKHRPSKPQARVAYESATAPKRHKKDPNRPRGYIS
AFNFFVKDKRPTYVQNRPNAQGHSLEHNNEINKILGKVWKMATEHEKKTYEEKATADKVR
YLKEMNAYRPPEGYERIAPRINVPKGYSSVAWDPDVNAGTSAQPCRRPWPAYTHFAHQEK
IGVLSSRSNRAKTLMSKAFGQRWHGMPDEETELYKELEEADKERYEKEAYAQDTAASLS

>KUK01repc678 (Schizocladia ischiensis)
MAGLYGLLAPKQLSGRLVVGVPANCGFERVPEDRFDKVVVETTLVTAGDIETRGWVGLSA
EGMLWVSQANVTSGMPADAVLRTLRTSGLQFLTGEGTALLSLQEFGQLESTLESCTILAF
DGNSAAAFAVMDGCENFCQDIEPGCYKLGISQPMDRKERQRFEEVALRMLAKGEHNVDTG
GIFSAFCDGTGNASSPPVFIHIIAGQDNTWGCHVHEQRGSMKVEVETLRTGSADPAIVIW
RHPRVIKDQGVLSWNQQDARQSNYMRTTTGDLSCCEYVVTNNGSSQLKLSISFAGNGKAS
TEMRRAPKRQKKDPNRPRGYISAFNFFVQIKRPHYLNQHPQVQDMSIEQNNVINKVLGRL
WKLATKDERRVYEEMATKDKLRYLREMKAYRPTQGYEREAPRINIPKGGDSKALALEFDT
PLAAMYLPRPKPAYTHFAHQDRRGVLSSGSVPAKALMSKTFAQRWQAMPSEEVTLYHELE
YADKERYYQDAKTHEDHTSTASGL

SRY + SOX Cluster

>sp|Q05066|SRY_HUMAN Sex-determining region Y protein OS=Homo sapiens OX=9606 GN=SRY PE=1 SV=1
MQSYASAMLSVFNSDDYSPAVQENIPALRRSSSFLCTESCNSKYQCETGENSKGNVQDRV
KRPMNAFIVWSRDQRRKMALENPRMRNSEISKQLGYQWKMLTEAEKWPFFQEAQKLQAMH
REKYPNYKYRPRRKAKMLPKNCSLLPADPASVLCSEVQLDNRLYRDDCTKATHSRMEHQL
GHLPPINAASSPQQRDRYSHWTKL

>sp|Q05738|SRY_MOUSE Sex-determining region Y protein OS=Mus musculus OX=10090 GN=Sry PE=1 SV=3
MEGHVKRPMNAFMVWSRGERHKLAQQNPSMQNTEISKQLGCRWKSLTEAEKRPFFQEAQR
LKILHREKYPNYKYQPHRRAKVSQRSGILQPAVASTKLYNLLQWDRNPHAITYRQDWSRA
AHLYSKNQQSFYWQPVDIPTGHLQQQQQQQQQQQFHNHHQQQQQFYDHHQQQQQQQQQQQ
QFHDHHQQKQQFHDHHQQQQQFHDHHHHHQEQQFHDHHQQQQQFHDHQQQQQQQQQQFH
DHHQQKQQFHDHHHHQQQQQFHDHQQQQQQFHDHQQQQHQFHDHPQQKQQFHDHPQQQQQ
FHDHHHQQQQKQQFHDHHQQKQQFHDHHQQKQQFHDHHQQQQQFHDHHQQQQQQQQQQQQ
QFHDQQLTYLLTADITGWKGIKHCTGPDPEPF

>tr|A0A0A7RQ35|A0A0A7RQ35_MONDO Sex determining region Y OS=Monodelphis domestica OX=13616 GN=SRY PE=4 SV=1
MYNFLEIKSSFVEEDLRVSESVKNNWDNRSGSISRVKRPMNAFMVWSRSQRRKVAQENPK
MHNSEISKLLGASWKLLTDNEKQPFIDEAKRLRAKHREEHPDYKYQPRRKTKSFMKNRQR
CYPKDRCTYGTSSLTQEQDTQKDLYSTTPQSYESNALISEISTFNYAQDPCTTHFGNWIN
VMNLPPEQENPEMWPLQNSGTVVNNIEHLTYI

>sp|P41225|SOX3_HUMAN Transcription factor SOX-3 OS=Homo sapiens OX=9606 GN=SOX3 PE=1 SV=2
MRPVRENSSGARSPRVPADLARSILISLPFPPDSLAHRPPSSAPTESQGLFTVAAPAPGA

PSPPATLAHLLPAPAMYSLLETELKNPVGTPTQAAGTGGPAAPGGAGKSSANAAGGANSG
GGSSGGASGGGGGTDQDRVKRPMNAFMVWSRGQRRKMALENPKMHNSEISKRLGADWKLL
TDAEKRPFIDEAKRLRAVHMKEYPDYKYRPRRKTKTLLKKDKYSLPSGLLPPGAAAAAAA
AAAAAAAASSPVGVGQRLDTYTHVNGWANGAYSLVQEQLGYAQPPSMSSPPPPPALPPMH
RYDMAGLQYSPMMPPGAQSYMNVAAAAAAASGYGGMAPSATAAAAAAYGQQPATAAAAAA
AAAAMSLGPMGSVVKSEPSSPPPAIASHSQRACLGDLRDMISMYLPPGGDAADAASPLPG
GRLHGVHQHYQGAGTAVNGTVPLTHI

>sp|P48436|SOX9_HUMAN Transcription factor SOX-9 OS=Homo sapiens OX=9606 GN=SOX9 PE=1 SV=1
MNLLDPFMKMTDEQEKGLSGAPSPTMSEDSAGSPCPSGSGSDTENTRPQENTFPKGEPDL
KKESEEDKFPVCIREAVSQVLKGYDWTLVPMPVRVNGSSKNKPHVKRPMNAFMVWAQAAR
RKLADQYPHLHNAELSKTLGKLWRLLNESEKRPFVEEAERLRVQHKKDHPDYKYQPRRRK
SVKNGQAEAEEATEQTHISPNAIFKALQADSPHSSSGMSEVHSPGEHSGQSQGPPTPPTT
PKTDVQPGKADLKREGRPLPEGGRQPPIDFRDVDIGELSSDVISNIETFDVNEFDQYLPP
NGHPGVPATHGQVTYTGSYGISSTAATPASAGHVWMSKQQAPPPPQQPPQAPPAPQAPP
QPQAAPPQQPAAPPQQPQAHTLTTLSSEPGQSQRTHIKTEQLSPSHYSEQQQHSPQQIAY
SPFNLPHYSPSYPPITRSQYDYTDHQNSSSYYSHAAGQGTGLYSTFTYMNPAQRPMYTPI
ADTSGVPSIPQTHSPQHWEQPVYTQLTRP

Transcription factor TCF/LEF family cluster

>sp|P36402|TCF7_HUMAN Transcription factor 7 OS=Homo sapiens OX=9606 GN=TCF7 PE=1 SV=3
MPQLDSGGGGAGGGDDLGAPDELLAFQDEGEEQDDKSRDSAAGPERDLAELKSSLVNESE
GAAGGAGIPGVPGAGAGARGEAEALGREHAAQRLFPDKLPEPLEDGLKAPECTSGMYKET
VYSAFNLLMHYPPPSGAGQHPQPQPPLHKANQPPHGVPQLSLYEHFNSPHPTPAPADISQ
KQVHRPLQTPDLSGFYSLTSGSMGQLPHTVSWFTHPSLMLGSGVPGHPAAIPHPAIVPPS
GKQELQPFDRNLKTQAESKAEKEAKKPTIKKPLNAFMLYMKEMRAKVIAECTLKESAAIN
QILGRRWHALSREEQAKYYELARKERQLHMQLYPGWSARDNYGKKKRRSREKHQESTTET
NWPRELKDGNGQESLSMSSSSSPA

>sp|Q8IMA8|PANG2_DROME Protein pangolin, isoform J OS=Drosophila melanogaster OX=7227 GN=pan PE=1
SV=2
MPHCGTTNSEINLNSNELISYQKKKSNEDLQKKHDKKCSINLKTSQVPDGGFNNQYTQLT
ANKIEQYNCNDLDKVCIVPSSRSDGMEASFSELLLSPNKLISQPWQTAEEIENWQNDSFR
QRNEMFSCIYTNSMLNQQPCSQQQLLATQLLYARLLRSQLAEREFHSNKFNMVHYSGSKK
TMLREDELLSTPSSQDNNNNIKLIKDIENSISCVDPPLFEFSNVHQRAEQKNKQDDKNCY
SPKLKSNKEALDGYDLQHTCDFIREQKNILIDIKKKLDNLSDSSGKFRKRLSVRQSHIEV
NNGSNSALEESVRRQLNGNRKSIENLLEEVKRLYNQWSSAELYYVRSLQRLGLPSEEDGE
YTSTPTHTIMALAAIALSNESILPQKTNAVHSKIPDVESESDFFLTSAKPKTLVEIEDII
LQLASSVNMHQSSSASTPHESYSDSVKSDCEESNSAPTCIWHSTRQTFRHKKDVEPCSTA
AEIILEYASLSSSSNIETSRLLTSASNLTDVTENYNVTTQLPIVFNYNRESSAESIVTDP
LLVPEFSTAPSTPSTGSNSGCSTGMVSGIFGLSQNRRKQRLARHIETLTTNSFKSNAIRG
NVDNEITNQLKTSPSIRALPTENVSHLITKTLSSANASLTAQELSITNLFKERLCALQGN
AGSMKTEIFPMIDAPYDLSIGSKTKHMNLEAKHTSNAQSNESKETTNDKKKPHIKKPLNA
FMLYMKEMRAKVVAECTLKESAAINQILGRRWHELSREEQSKYYEKARQERQLHMELYPG
WSARDNYGYVSKKKKRKKDRSTTDSGGNNMKKCRARFGLDQQSQWCKPCRRKKKCIRYME
ALNGNGPAEDGSCFDEHGSQLSDDDEDDYDDDKLGGSCGSADETNKIEDEDSESLNQSMP
SPGCLSGLSSLQSPSTTMSLASPLNMNANSATNVIFPASSNALLIVGADQPTAQQRPTLV
STSGSSSGSTSSISTTPNTSSTVSPVTCMTGPCLGSSQERAMMLGNRFSHLGMGLSPPVV
STSTSKSEPFFKPHPTVCNNPIFALPSIGNCSLNISSMPNTSRNPIGANPRDINNPLSIN
QLTKRREYKNVELIEASESKTIVAHAATSIIQHVAVNGYHANHSLLNSNLGHLHHQLNNR
TENPNRSEQTMLSVSNHSVNSSECHKESDSQAIVSSNPPNAGSSDNGVISVS

>tr|B4DZY5|B4DZY5_HUMAN cDNA FLJ54043, highly similar to Lymphoid enhancer-binding factor 1 OS=Homo
sapiens OX=9606 PE=2 SV=1
MMPNMNNDPYMSNGSLSPPIPRTSNKVPVVQPSHAVHPLTPLITYSDEHFSPGSHPSHIP
SDVNSKQGMSRHPPAPDIPTFYPLSPGGVGQITPPLGWFSHHMIPGPPGPHTTGIPHPAI
VTPQVKQEHPHTDSDLMHVKPQHEQRKEQEPKRPHIKKPLNAFMLYMKEMRANVVAECTL
KESAAINQILGRRWHALSREEQAKYYELARKERQLHMQLYPGWSARDNYGKKKRKREKL
QESASGTGPRMTAAYI

TOX family cluster

>sp|O94900|TOX_HUMAN Thymocyte selection-associated high mobility group box protein TOX OS=Homo sapiens OX=9606 GN=TOX PE=1 SV=3
MDVRFYPPPAQPAAAPDAPCLGPSPCLDPYYCNKFDGENMYMSMTEPSQDYVPASQSYPG
PSLESEDFNIPPITPPSLPDHSLVHLNEVESGYHSLCHPMNHNGLLPFHPQNMDLPEITV
SNMLGQDGTLLSNSISVMPDIRNPEGTQYSSHPQMAAMRPRGQPADIRQQPGMMPHGQLT
TINQSQLSAQLGLNMGGSNVPHNSPSPPGSKSATPSPSSSVHEDEGDDTSKINGGEKRPA
SDMGKKPKTPKKKKKKDPNEPQKPVSAYALFFRDTQAAIKGQNPNATFGEVSKIVASMWD
GLGEEQKQVYKKKTEAAKKEYLKQLAAYRASLVSKSYSEPVDVKTSQPPQLINSKPSVFH
GPSQAHSALYLSSHYHQQPGMNPHLTAMHPSLPRNIAPKPNNQMPVTVSIANMAVSPPPP
LQISPPLHQHLNMQQHQPLTMQQPLGNQLPMQVQSALHSPTMQQGFTLQPDYQTIINPTS
TAAQVVTQAMEYVRSGCRNPPPQPVDWNNDYCSSGGMQRDKALYLT

Nucleolar transcription factor 1 cluster

>sp|P17480|UBF1_HUMAN Nucleolar transcription factor 1 OS=Homo sapiens OX=9606 GN=UBTF PE=1 SV=1
MNGEADCPTDLEMAAPKGQDRWSQEDMLTLLECMKNNLPSNDSSKFKTTESHMDWEKVAF
KDFSGDMCKLKWVEISNEVRKFRTLTELILDAQEHVKNPYKGKKLKKHPDFPKKPLTPYF
RFFMEKRAKYAKLHPEMSNLDLTKILSKKYKELPEKKKMKYIQDFQREKQEFERNLARFR
EDHPDLIQNAKKSDIPEKPKTPQQLWYTHEKKVYLKVRPDATTKEVKDSLGKQWSQLSDK
KRLKWIHKALEQRKEYEEIMRDYIQKHPELNISEEGITKSTLTKAERQLKDKFDGRPTKP
PPNSYSLYCAELMANMKDVPSTERMVLCSQQWKLLSQKEKDAYHKKCDQKKKDYEVELLR
FLESLPEEEQQRVLGEEKMLNINKKQATSPASKKPAQEGGKGGSEKPKRPVSAMFIFSEE
KRRQLQEERPELSESELTRLLARMWNDLSEKKKAKYKAREAALKAQSERKPGGEREERGK
LPESPKRAEEIWQQSVIGDYLARFKNDRVKALKAMEMTWNNMEKKEKLMWIKKAAEDQKR
YERELSEMRAPPAATNSSKKMKFQGEPKKPPMNGYQKFSQELLSNGELNHLPLKERMVEI
GSRWQRISQSQKEHYKKLAEEQQKQYKVHLDLWVKSLSPQDRAAYKEYISNKRKSMTKLR
GPNPKSSRTTLQSKSESEEDDEEDEDDEDEDEEEEDDENGDSSEDGGDSSESSSEDESED
GDENEEDDEDEDDDEDDDEDEDNESEGSSSSSSSSGDSSDSDSN

HMGB family cluster

>sp|P09429|HMGB1_HUMAN High mobility group protein B1 OS=Homo sapiens OX=9606 GN=HMGB1 PE=1 SV=3
MGKGDPKKPRGKMSSYAFFVQTCREEHKKKHPDASVNFSEFSKKCSERWKTMSAKEKGKF
EDMAKADKARYEREMKTYIPPKGETKKKFKDPNAPKRPPSAFFLFCSEYRPKIKGEHPGL
SIGDVAKKLGEMWNNTAADDKQPYEKKAAKLKEKYEKDIAAYRAKGKPDAAKKGVVKAEK
SKKKKEEEEDEEDEEDEEEEEDEEDEDEEEDDDDE

>sp|P26583|HMGB2_HUMAN High mobility group protein B2 OS=Homo sapiens OX=9606 GN=HMGB2 PE=1 SV=2
MGKGDPNKPRGKMSSYAFFVQTCREEHKKKHPDSSVNFAEFSKKCSERWKTMSAKEKSKF
EDMAKSDKARYDREMKNYVPPKGDKKGKKKDPNAPKRPPSAFFLFCSEHRPKIKSEHPGL
SIGDTAKKLGEMWSEQSAKDKQPYEQKAAKLKEKYEKDIAAYRAKGKSEAGKKGPGRPTG
SKKKNEPEDEEEEEEEEDEDEEEEDEDEE

FACT-SSRP1 cluster

>sp|Q08945|SSRP1_HUMAN FACT complex subunit SSRP1 OS=Homo sapiens OX=9606 GN=SSRP1 PE=1 SV=1
MAETLEFNDVYQEVKGSMNDGRLRLSRQGIIFKNSKTGKVDNIQAGELTEGIWRRVALGH
GLKLLTKNGHVYKYDGFRESEFEKLSDFFKTHYRLELMEKDLCVKGWNWGTVKFGGQLLS
FDIGDQPVFEIPLSNVSQCTTGKNEVTLEFHQNDDAEVSLMEVRFYVPPTQEDGVDPVEA
FAQNVLSKADVIQATGDAICIFRELQCLTPRGRYDIRIYPTFLHLHGKTFDYKIPYTTVL
RLFLLPHKDQRQMFFVISLDPPIKQGQTRYHFLILLFSKDEDISLTLNMNEEEVEKRFEG
RLTKNMSGSLYEMVSRVMKALVNRKITVPGNFQGHSGAQCITCSYKASSGLLYPLERGFI
YVHKPPVHIRFDEISFVNFARGTTTTRSFDFEIETKQGTQYTFSSIEREEYGKLFDFVNA
KKLNIKNRGLKEGMNPSYDEYADSDEDQHDAYLERMKEEGKIREENANDSSDDSGEETDE
SFNPGEEEEDVAEEFDSNASASSSSNEGDSDRDEKKRKQLKKAKMAKDRKSRKKPVEVKK
GKDPNAPKRPMSAYMLWLNASREKIKSDHPGISITDLSKKAGEIWKGMSKEKKEEWDRKA
EDARRDYEKAMKEYEGGRGESSKRDKSKKKKKVKVKMEKKSTPSRGSSSKSSSRQLSESF
KSKEFVSSDESSSGENKSKKKRRRSEDSEEEELASTPPSSEDSASGSDE

Non-histone chromosomal protein 6 cluster

>sp|P11632|NHP6A_YEAST Non-histone chromosomal protein 6A OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=NHP6A PE=1 SV=1
MVTPREPKKRTTRKKKDPNAPKRALSAYMFFANENRDIVRSENPDITFGQVGKKLGEKWK
ALTPEEKQPYEAKAQADKKRYESEKELYNATLA

SexP cluster

>sexP tr|B0F2H1|B0F2H1_PHYBL Sex plus OS=Phycomyces blakesleeanus OX=4837 GN=sexP PE=4 SV=1
MKAKQTPHLPYQRPIKPRLLLAPKLSVQQISVIYIKGKTALRIYPSQAFVYARDLITMVD
EKRDYITVNTGSSVERKDLFSAAESLLRKQMCTESSSTGVSISKGQENPYQASPEQVPKR
PSNAFMIYSATLRKRIKTTFPEYNNSDISKLLGAMWKNAGAEVKKEYMEKANEVREWHKE
RYPDYEYNSRKQSTKERDSMPRTDFSNQDFITADDEWIRQLNDLLSQNPMTALHGSSQEN
FAIDYTSTPYFISSPSNDISLGDLKLEEWQDFCNNWTEHDSRYLDQNFLDESFWARF

SexM cluster

>tr|A0A167RE73|A0A167RE73_PHYB8 HMG box domain-containing protein OS=Phycomyces blakesleeanus (strain ATCC 8743b / DSM 1359 / FGSC 10004 / NBRC 33097 / NRRL 1555) OX=763407 GN=PHYBLDRAFT_154054 PE=4 SV=1
MESFMFVNEFKNSYQPPSDILTESISSTPVRQKCIISSTVKIRRPKNAFMLYRQAVHPSI
LSSNSTIHNKEISRTAGKMWKNEKEEVRKYYERKADEEKLYHSKKFPGYIYKPQQRKTRR
PQSTVCKPFLRSTSDIQLIYQKTESKSPSEKLSATSQDYIIHCRDESTNEDEILNTPPTQ
CLDDSILKTFLDDLSFDIFSPIDELIFSLCS

>sp|P19392|MATA_NEUCR Mating-type protein A-1 OS=Neurospora crassa (strain ATCC 24698 / 74-OR23-1A / CBS 708.71 / DSM 1257 / FGSC 987) OX=367110 GN=mtA-1 PE=1 SV=2
MSGVDQIVKTFADLAEDDREAAMRAFSRMMRRGTEPVRRIPAAKKKVNGFMGFRSYYSPL
FSQLPQKERSPFMTILWQHDPFHNEWDFMCSVYSSIRTYLEQEKVTLQLWIHYAVGHLGV
IIRDNYMASFGWNLVRFPNGTHDLERTALPLVQHNLQPMNGLCLLTKCLESGLPLANPHS
VIAKLSDPSYDMIWFNKRPHRQQGHAVQTDESEVGVSAMFPRNHTVAAEVDGIINLPLSH
WIQQGEFGTESGYSAQFETLLDSILENGHASSNDPYNMALAIDVPMMGFNGGA

>sp|P0CY17|MATMC_SCHPO Mating-type M-specific polypeptide Mc OS=Schizosaccharomyces pombe (strain 972 / ATCC 24843) OX=284812 GN=mat1-Mc PE=2 SV=1
MDSHQELSAGSPISYDFLDPDWCFKRYLTKDALHSIETGKGAAYFVPDGFTPILIPNSQS
YLLDGNSAQLPRPQPISFTLDQCKVPGYILKSLRKDTTSTERTPRPPNAFILYRKEKHAT
LLKSNPSINNSQVSKLVGEMWRNESKEVRMRYFKMSEFYKAQHQKMYPGYKYQPRKNKVK
R

>tr|Q7Z8M2|Q7Z8M2_EMEND MAT2 protein OS=Emericella nidulans OX=162425 GN=matA-2 PE=2 SV=1
MAAVSIAMKSPTQSPDSITELLWKDALRHLGSTNDEVLLPTNVVDIIGQDNVEKIKSRLS
ALLGAPVVSFVDESINALRVLRTPTFSGSSISVASPSRALDSWPSEPPNKPRPASMKPAK
IPRPPNAFILYRQHHYPKVKEARPDLSNNEISVIIGKKWRAEPEEGKLHFKNLAEEFKKK
HAEEYPDYQYTPRKPSEKKRRAASRISPKNSKRTVALENPGSMTAPSSNVFTPQMYPGIQ
NGQLAGAGYIGYLDGLNSMVNTGGLTDEPTNFGTNAFNSLFQQPQSDYGRTALFPQLEFA
GPSLGDSLEFPEFAADYF

Mating-type cluster 1

>sp|Q71U11|MATA2_CANAL Mating-type-like protein A2 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=MTLA2 PE=3 SV=1
MPYTFNFPKSQSSFKYSIKSLTNSRSFPMISIYVTPQSTPKFINNQSIDKMTVLSSFRSR
NSFIIARSILSKLLRKNKADFKKVSKSVSLLWGSVDNSFKNYFEYLSILESQWHDKKSSN
LNPISSTRTTSELNYIELCSSIYNQSRKHPRSVQGNKAKFRLYKNVSKKSNRKPRRLKNG
FKVSRSTINCSYGIYTEDVFQ

>sp|Q9UW19|MTAL1_CANAL Mating-type-like protein ALPHA1 OS=Candida albicans (strain SC5314 / ATCC MYA-2876) OX=237561 GN=MTLALPHA1 PE=2 SV=1
MGNKKKTRKTVPKEFISLFRVHSGRDAPRRDTREVQKSKKHGFRFTSLPDLPVASNALQE
LLLEYGLLNDIKWDSKGLKASKNKKTKLKPINSFIAFRSFYSRTISNPEHQRELSSKLAD
VWTQESNQEVWKQYTQSYNNYLLLPDAKLNFVDWLCEALDYTIDNTTPQIEDISLTSYNQ
LLSGTIEDVYIMK

Mating-type cluster 2

>sp|P0CY06|MTAL1_YEAST Mating-type protein ALPHA1 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=MATALPHA1 PE=1 SV=1
MFTSKPAFKIKNKASKSYRNTAVSKKLKEKRLAEHVRPSCFNIIRPLKKDIQIPVPSSRF
LNKIQIHRIASGSQNTQFRQFNKTSIKSSKKYLNSFMAFRAYYSQFGSGVKQNVLSSLLA
EEWHADKMQHGIWDYFAQQYNFINPGFGFVEWLTNNYAEVRGDGYWEDVFVHLAL

Mating-type cluster 3

>tr|G5EAT5|G5EAT5_EMENI Alpha box domain-containing protein; matB; Alpha box OS=Emericella nidulans (strain FGSC A4 / ATCC 38163 / CBS 112.46 / NRRL 194 / M139) OX=227321 GN=MAT1 PE=3 SV=1
MENALSPLQRAFNAFLLSMPPQQLDDLVKHIQDVKAQEQKPPVFRNEIPAIRANTTQDAH
HTFPTFPSSKHRPASSRGRRVHDGKRRPLNSFIAFRSFYSAIFPDITQKSKSGILRFLWQ
NDPFKAKWTILAKAYSIIRDKHDDEVSLESFLTLNAELIGVTQPDRYLDAMGWELTLNDQ
QQYTMARVKSPVATEAQLSTHFSVDDLIKHCYATGYVTEDKRKKEIRGHNAPVMTFATQP
ALVIHKNNSLQISGNHTVVSTNGSESVTKETPAFEPTEATELPYPSDIVSPVTGDTSFES
TDATRIYQRPQSRTSLAENYLDMANMQFHTWDDQTALLPYNTGPLMQESFDALDFKPFLN
I

>sp|P19392|MATA_NEUCR Mating-type protein A-1 OS=Neurospora crassa (strain ATCC 24698 / 74-OR23-1A / CBS 708.71 / DSM 1257 / FGSC 987) OX=367110 GN=mtA-1 PE=1 SV=2
MSGVDQIVKTFADLAEDDREAAMRAFSRMMRRGTEPVRRIPAAKKKVNGFMGFRSYYSPL
FSQLPQKERSPFMTILWQHDPFHNEWDFMCSVYSSIRTYLEQEKVTLQLWIHYAVGHLGV
IIRDNYMASFGWNLVRFPNGTHDLERTALPLVQHNLQPMNGLCLLTKCLESGLPLANPHS
VIAKLSDPSYDMIWFNKRPHRQQGHAVQTDESEVGVSAMFPRNHTVAAEVDGIINLPLSH
WIQQGEFGTESGYSAQFETLLDSILENGHASSNDPYNMALAIDVPMMGFNGGA

Mating-type cluster 4

>sp|Q10116|MATD_NEUCR Mating-type protein A-3 OS=Neurospora crassa (strain ATCC 24698 / 74-OR23-1A / CBS 708.71 / DSM 1257 / FGSC 987) OX=367110 GN=mtA-3 PE=2 SV=1
MSALDVDSISDIAPGLSPVTAIHYGRIQVMLFRSHLAEFAEEDLVYAMDNSVVVFGEEAL
LMVAPDESSIAICTYPFGLMMMEWGNWDILAVSPPSRTPTIPSESVLGISNQGGANVEQQ
EQSSHTIDMTLPSNFFEQSSVTQSNGTSRPRNQFVLYYQWLLDTLFSEDPSLSARNISQI
VAGLWNSEHPAAKARFRELAEMEVHRHRAENPHLYPDQPRFPTTDPVPPRMRYPCVISPE
DRQRILRMLDFVWEESNGQLAAEEAALNDVVQPQQAEEVGPFPDFEWEEPNHIIDMSTDL
SVAQDPDFMMTEDDSMRFLLKQAS

Closely-related proteins to HMG-sex from EukProt (Fig. 5B)

>EP00908_Phaeocystis_cordata_P015613
TQMVARPPTAAATQYVHGTIKVGADGRSKWRVDDAEYYTSQGVRGVARHAPNKGACEWTL
VERGDAPTESNASLVAARAAARAAARATAAQGGGGGGGAKMSAGSAEPAAVKTEARPAAA
ASLVEPRRPQNAYLLFAGATRAAVKADHPHLSLGKVGKLLGGRYQELSEEEKAHWQAEAE
ADKLRYEAELAEFRR

>EP00631_Chrysoreinhardia_giraudii_P005316 CAMPEP_0197417442
HRPRGYVSAFNFFARDCRRLMLEQNPTLQVASTDQNNSINKTLGLRWKRLSQEQRGTYEE
AAKLDKLRYLKEMEAYEPSVGFSKDPPRINPPRGFNCDGTPKQLHSSCRSQSQPPMVYRP
ISAYTSFVQHESTYLSAKAASHYLLSPRSRYFSERW

>EP00633_Aureoumbra_lagunensis_P000885 CAMPEP_0197288952
MTSENAALDNEFMFNYLFQVLTEFHKSTSTLPLSSFRSSWFCTDCQLIYGIQKRVCHVIF
RNNSVITSKQLRHGIHFIESFGSLNKSNLNQQHAPRWLCGEGTTLARLTIRADFEKAILT
DEYFDETSGFWNNMTKQFRYDTCLDVVTGVNHSHLGRRRKDPNKPRGYISAFNYFAKSFR
NGLGDQNPGLDIRSPEQNNRLNRITGRRWKLLLSSEREYYDIAAVRDKIRYLKEMKAYQP
PVGHSKCLPRINPPSGIDWDGTVVNSSGDATVVNTTGRPRPISEYTVFVQQEIMDIGVKC
ANLSNHSRYFGMRWKSMTQDEHRLYEYLADKMNKRHM

>EP00617_Octactis_speculum_P025241 CAMPEP_0185750782
XGIMITDFIAATLTQNEANGNPVSIGEAHDILGSCLENFLQDDAAIDDVLKQRSSLPLCL
ANHEYDLILSHHKFGLLWFSKDGEICVTNKLKIQEIPMKFQGPNDPPISLNVHRLFDMMT

AKANEVGRVSPTRSKSGRSAVHKKKTAIQCIQLIFPGPEILHGHREVHGCSEEIAFRYIL
NGKQTNKNTKRKTRDPNLPAGYISGFNFFAKDCRPRLLREYPELKVHGATNNEINKLIGK
LWKSLPLKFRAEFEQRSWHDKLRYFEEMKQYSPTDGFKRQIPRIHPVDSPSSFKNGFTGK
KIKTYTSNNTGDNRISQSGYSVFVQQECSQLAPQMNAKYTPRDATSDLSFVLVERWKSMR
DDEKHLYDN