Current Biology

Insights into the Evolution of Multicellularity from the Sea Lettuce Genome

--Manuscript Draft--

Manuscript Number:	CURRENT-BIOLOGY-D-18-00475R1
Full Title:	Insights into the Evolution of Multicellularity from the Sea Lettuce Genome
Article Type:	Research Article
Corresponding Author:	Olivier De Clerck Ghent University Ghent, BELGIUM
First Author:	Olivier De Clerck
Order of Authors:	Olivier De Clerck
	Shu-Min Kao
	Kenny A. Bogaert
	Jonas Blomme
	Fatima Foflonker
	Michiel Kwantes
	Emmelien Vancaester
	Lisa Vanderstraeten
	Eylem Aydogdu
	Jens Boesger
	Gianmaria Califano
	Benedicte Charrier
	Rachel Clewes
	Andrea Del Cortona
	Sofie D'Hondt
	Noe Fernandez-Pozo
	Claire M. Gachon
	Marc Hanikenne
	Linda Lattermann
	Frederik Leliaert
	Xiaojie Liu
	Christine A. Maggs
	Zoe A. Popper
	John A. Raven
	Michiel Van Bel
	Per K.I. Wilhelmsson
	Debashish Bhattacharya
	Juliet C. Coates
	Stefan A. Rensing
	Dominique Van Der Straeten

	Assaf Vardi
	Lieven Sterck
	Klaas Vandepoele
	Yves Van de Peer
	Thomas Wichard
	John H. Bothwell
Abstract:	We report here the 98.5 Mbp haploid genome (12,924 protein coding genes) of Ulva mutabilis, a ubiquitous and iconic representative of the Ulvophyceae or green seaweeds. Ulva's rapid and abundant growth makes it a key contributor to coastal biogeochemical cycles; its role in marine sulfur cycles is particularly important because it produces high levels of dimethylsulfoniopropionate (DMSP), the main precursor of volatile dimethyl sulfide (DMS). Rapid growth makes Ulva attractive biomass feedstock, but also increasingly a driver of nuisance 'green tides'. Additionally, ulvophytes are key to understanding evolution of multicellularity in the green lineage. Furthermore, morphogenesis is dependent on bacterial signals, making it an important species to study cross-kingdom communication. Our sequenced genome informs these aspects of ulvophyte cell biology, physiology and ecology. Gene family expansions associated with multicellularity are distinct from those of freshwater algae. Candidate genes are present for the transport and metabolism of DMSP, including some that arose following horizontal gene transfer from chromalveolates. The Ulva genome offers, therefore, new opportunities to understand coastal and marine ecosystems, and the fundamental evolution of the green lineage.

Dear Editor,
Dear Dr. North,

Please find attached a revised version of our manuscript reporting the Ulva genome (CBIOL D-18-00475). We would like to thank both reviewers for their constructive criticism. We have incorporated virtually all comments by the reviewers. In a separate rebuttal letter we detail point-by-point how and where we adapted the manuscript. We have also made the text conform the format of Current Biology (e.g. Star Methods, and the Supplemental Experimental Procedures are incorporated in the main text).

We feel confident that we have addressed the reviewer comments satisfactorily and have included a full rebuttal with the submission of our manuscript. Should there be additional questions, we would be very happy to respond to these.

One reviewer commented that he could not access the sequence data and gene models. Thereto, we have provided a reviewer login to the Orcae system where all relevant genome data can be found.

- http://bioinformatics.psb.ugent.be/orcae/users/login
    - Username: ulvmu_reviewer
    - Password: 1800475
- The assembly can be downloaded at https://bioinformatics.psb.ugent.be/gdb/ulva/Ulvmu_genome_LATEST.tfa.gz using the same credential

From a technical perspective, the main text is 4491 words with 7 figures and 1 table. Supplementary information consists of 5 figures and 9 tables.

Yours sincerely,

Olivier De Clerck
[on behalf of the *Ulva* genome consortium]

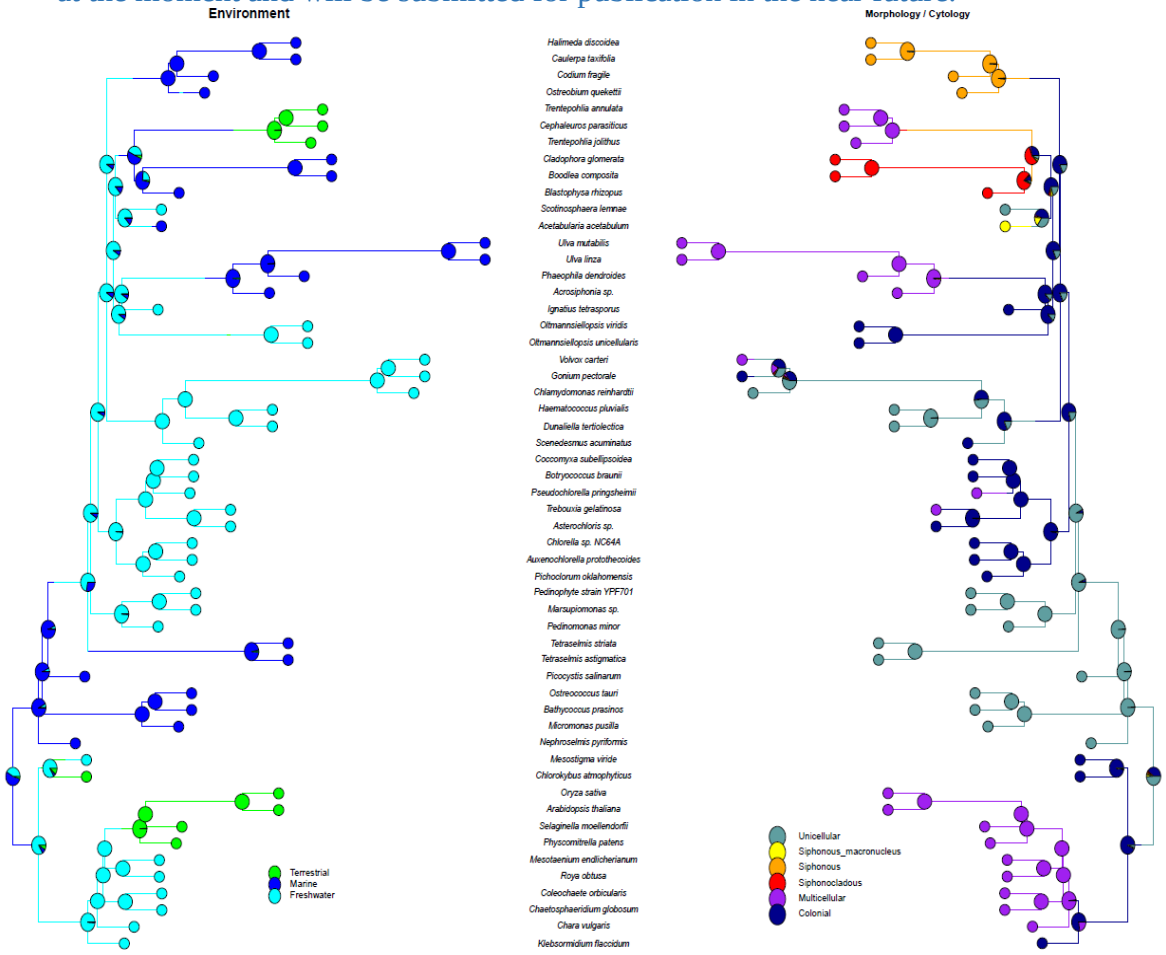
Reviewer #1: This manuscript describes the genome sequence and its characterization from Ulva mutabilis, and is the first genome from the Ulvophyceae, one of three major green algal crown groups. The phylogenetic position of Ulva makes this genome paper an important addition to our understanding of green algal genomes and evolution. Ulva mutabilis is also a multicellular species and has some interesting biotic and metabolic interactions that add interest and significance to the results. The paper is well written and the figures are generally well constructed. I have several suggestions for the authors to improve the presentation. Note also there were several items that I could not adequately review because access credentials for the Ulva genome assembly and browser were not provided.

> We have provided a reviewer login to the Orcae system where all relevant genome data can be found.
>
> http://bioinformatics.psb.ugent.be/orcae/users/login
>
> Username: ulvmu_reviewer
>
> Password: 1800475The assembly can be downloaded at
>
> https://bioinformatics.psb.ugent.be/gdb/ulva/Ulvmu_genome_LATEST.tfa.gz
>
> using the same credential

1. Genomic resources for Ulvophytes are relatively scarce but not completely absent, and I feel the authors should have taken advantage of what is available to strengthen and help contextualize their results. First, there is a large transcriptome data set from Caulerpa taxifolia (10.1371/journal.pgen.1004900.s031) which includes a file of all predicted CDS that could be translated and used for comparison with the predicted Ulva proteome (or expressed proteome) to help assess conservation and divergence of Ulvophyte protein coding genes. The second resource is the mating type locus assemblies of U. partita , a close relative of U. mutabilis, which could be used to compare mating type region divergence/conservation between the minus haplotype of U. mutabilis (sequenced in this manuscript) and its homolog from U. partita.

> Indeed we had initially overlooked a comparison of the *Ulva* genome with genomic resources from other Ulvophyceae (e.g. *Caulerpa*, *Ulva partita*). Thereto, we checked whether the most important findings with respect to gains and expansions (e.g. CONSTANS-like expansion, receptor kinases, pectin lyases) are mirrored in the *Caulerpa* genomic resources. We mention these studies now in the Introduction and incorporated a comparison in the relevant sections of text.
>
> With respect to the findings, the expansions in the *Ulva* genome are not consistently reflected in the *Caulerpa* transcriptome. For example, COLS and SRCR are absent in the latter. A gain of the pectin lyase fold domain is shared between *Ulva* and *Caulerpa*, however, in the latter there is no coupling with kinase domains. It is important to realise that these results have to be interpreted with caution giventhat for *Caulerpa* only transcriptome data are available.
>
> A side note but worth mentioning in this context, the large differences between *Caulerpa* and *Ulva* are not unexpected. As part of an ongoing study we produced a phylotranscriptomic analyses of the Chlorophyta (55 taxa x 600 genes) which highlights a

radiation of the Ulvophyceae and Chlorophyceae (~600-700 mya). In other words *Ulva* is not closer related to the Chlorophyceae (*Chlamydomonas*, *Volvox*, etc.) than it is to the siphonous seaweeds (e.g. *Caulerpa*). Furthermore, nearly every transition to the marine environment seems to have coincided with a different solution for macroscopic growth (siphonous, siphonocladous, multicellular, etc.). We refer to the phylogeny and ancestral reconstruction below. We are very excited about these results, which are being analyzed at the moment and will be submitted for publication in the near future.



With respect to *Ulva partita* the genomic information has not been released and only the genes of the mating type locus have deposited in GenBank. We checked whether mating type specific genes are resolved on a single contig in *Ulva mutabilis* which is not the case. Results however are difficult to interpret at this stage. Both mt- and mt+ mating type locus genes are in the *U. mutabilis* wildtype genome (on 7 contigs). In other words, the mating type specific genes identified in *U. partita* do not seem to sort according to mating type in *U. mutabilis*. We did not include this information that hints toward a dynamic evolution of mating loci.

2. It was somewhat disappointing that the authors did not incorporate more data on biotic interactions with bacteria that are the most exciting aspect of multicellularity in Ulva. Transcriptome comparisons of xenic versus axenic cultures would have added a large amount of interest to an otherwise fairly standard genome paper. Inclusion of even a more limited analysis

of expression for DMSP related genes (Fig. 6) or iron related genes (Table S6) in the presence/absence of the two bacterial partners used for co-culture would have been nice. The slender mutant and its impact on differentiation was another interesting avenue for investigation that was not fully exploited. I do understand that genome papers are necessarily limited in scope, but more and more often these days there is at least some additional experimental data to help elevate a genome paper for general interest journals.

> We agree that the reliance of Ulva on bacteria is very interesting. We did indeed try to obtain a better understanding of the algal-bacterial interactions by looking at horizontal gene transfer, transcription DMSP-related genes, phytohormones (both xenic and axenic conditions). The last three have been complemented with experimental evidence and RNA-seq libraries in axenic and non-axenic conditions.

3. The section on plant hormones was moderately interesting, but essentially there was no evidence found for any of the detected hormones having a strong function in Ulva development. I suggest moving Table 2 to the Supplement, and replacing it with a graphic/table summarizing which hormonal pathways/genes have been detected across the chlorophyte lineage, including Ulva. This would help contextualize the results from Ulva by showing whether it is exceptional in some way with respect to its phytohormone related gene repertoire compared with unicellular chlorophytes. Another reason to add such a Table/Figure is that the Wang reference (62) does not include a survey of phytohormone biosynthetic genes, and additional genome data are now available from several chlorophyte groups that were not available in 2015.

> Table 2 has been transferred to the Supplement and has been replaced by a comprehensive overview of the phytohormone synthesis in green algae. The value of our measurements is situated in the xenic versus axenic conditions. Reports of phytohormones in algae are very common (e.g. Stirk & Van Staden 2014, Adv. Bot. Res. 71:125-59) but it is never clear whether the phytohormones are produced by the algae themselves are associated bacteria. Now that we demonstrate that *Ulva* produces most common phytohormones (axenic and xenic conditions), further research can focus on putative functions. We stressed the latter in the revised version of the manuscript..

Additional points.

4. The analysis of HGT was interesting and the bioinformatics portion of the discovery pipeline seemed thorough, but I did not find any description in the Methods about manual validation for the genes in Table S5. Relying on purely in silico methods can be misleading. For example, were any of these 22 HGT genes found on orphan scaffolds or in poorly assembled genomic regions (both red flags that might indicate contamination rather than HGT)?
Did the HGT genes look like Ulva genes in terms of GC content and structure (presence of introns)? If the authors have done this level of validation then it should be mentioned in the Methods (and I apologize if I missed it). Also would be useful to know for each gene reported in Table S5: (i) whether there is evidence for expression (from transcriptome data), (ii) whether the gene has a any vertically inherited paralogs in Ulva (with the HGT copy being an extra copy), (iii) whether the HGT-acquired gene might have displaced or substituted for an ancestral

homolog that had been vertically inherited prior to the emergence of the Ulva lineage (inferred from looking at homologs in other chlorophyte algae), (iv) what is the most similar homolog for each gene.

> We agree that prokaryote to eukaryote HGT is a vividly discussed nowadays and we were happy to conduct these additional checks. As it turned out we discovered a mistake in our pipeline for which we apologise. Apparently, in our original implementation, duplicated genes were not showing up in the analyses. Rerunning the analyses resulted in a very interesting picture whereby more than half of the HGT-genes seemed duplicated at least once. In one instance (Haem peroxidases) we discovered 36 copies resulting from a single HGT event. We have added this information to the text. In addition we adapted Table S5 to include expression values (FPKM), number of introns, GC content. In summary, most genes are expressed, all but one gene also have introns.
>
> Non-HGT paralogs and displacement of ancestral homologs are difficult to point. By their nature these putative HGT genes are highly divergent from the remainder of the gene families. In our gene families HGT genes are resolved in their own families, without any genes from other green algae.
>
> Lastly, we confirm that the HGT genes were recovered as part of unsuspicious sequencing reads (i.e. being flank up and or downstream by non-HGT genes) and included this information in the material and methods. The reviewer may want to consult a visual digest at https://bioinformatics.psb.ugent.be/gdb/ulva/HGT/mapping.zip

5. In Fig. 2B the graph could be more informative with a couple changes. (i) will be helpful if the data are presented in a normalized frequency spectrum (i.e. % of genes rather than absolute numbers). (ii) the fraction of singletons (non-duplicated genes) should be listed in the legend or shown as a separate bar graph with different y axis. (iii) A finer-scale breakdown of data that includes the (2-5) duplicate copy number category should be included.

> Normalizing over number of gene families or gene makes the trend disappear. In other words, the initial graph was misleading in that the larger gene families sizes were due to a higher number of genes. The text has been adapted accordingly and the graph removed. An adapted figure in the supplement illustrates this trend,

6. There was some important and interesting background information in the Supplement on the cell cycle relate genes (lines 291-301) that should be moved to the Main text because they are of general interest for understanding multicellularity in Ulva.

> We added this information to the main text of the revised version.

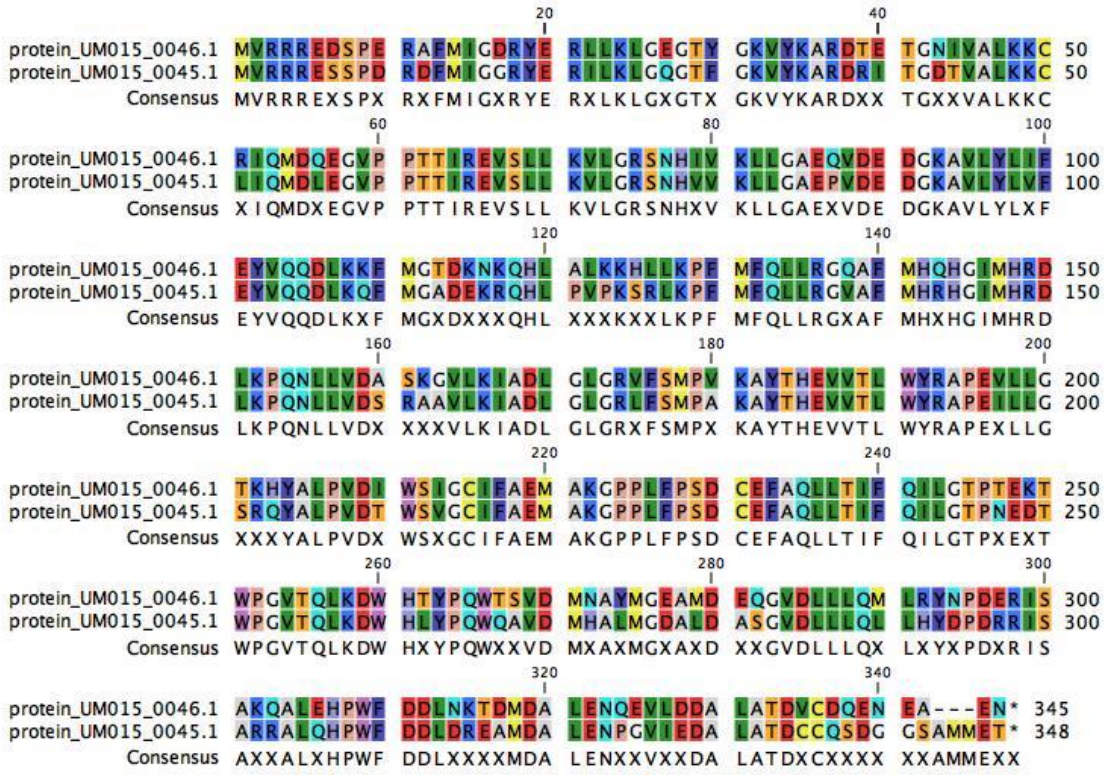
7. The loss of RB pathway was really interesting and is a convincing finding since it appears to involve the entire set of genes and not just absence of one. The authors could also query the Caulerpa data to determine whether the loss of this pathway might be early in the history of Ulvophyceae or was possibly taxon specific.

Retinoblastoma and E2F are both present in the *Caulerpa* transcriptomic dataset (reciprocal blast hit). We added this information to the main text.

The loss of a major conserved cell cycle pathway in Ulva makes it even more important to adequately characterize the other predicted cell cycle regulators encoded in its genome. There were some additional questions raised for me in the Supplemental Table of cell cycle regulators. (i) The CDKA ortholog is described as being non-canonical in its cyclin interacting motif that is normally PSTAIRE for this sub-group. The authors should really make (and present) data for a phylogeny to explore this finding that will either support or refute the inclusion of their candidate CDKA in the CDK1 clade or place it in other CDK sub-clades. (ii) Two gene models are shown for CDKB but this potential duplication is not mentioned. Is it a correct result? (iii) Two models for B-type cyclins and some ambiguous annotation surrounding another cyclin noted as ?? AB suggest that these relationships need to be worked out using a phylogenetic approach. (iv) Did the authors search for a CDC25 ortholog? This gene is not universally present in green algae but has been found in prasinophytes. (v) Please use some consistent notation in the Tables like ND for not detected. Not sure what NA means.

Additional information on the divergent CDKA gene is now incorporated in the main text and a phylogenetic tree showing its position close to other green algal CDKAs is added as a supplementary figure.

CDKB is indeed duplicated in Ulva (UM015_0045 and UM015_0046; see Suppl. Data1). As shown below they are a tandem repeat but are not exact matches. We have no expression values of UM01_0045, which may point toward a pseudogene. We have not included the information because of its speculative nature.



We have provided a phylogenetic tree for the cyclins in Supplement for the Cyclin. The ambiguously annotated gene model turned out to be a Cyclin B.

No CDC25 ortholog was detected, using *Chlamydomonas* [XP_001702069] and *Arabidopsis* (AT5G03455) as a blast query.

Inconsistent in Suppl.Data1.xls annotation has been replaced by 'Not detected'.

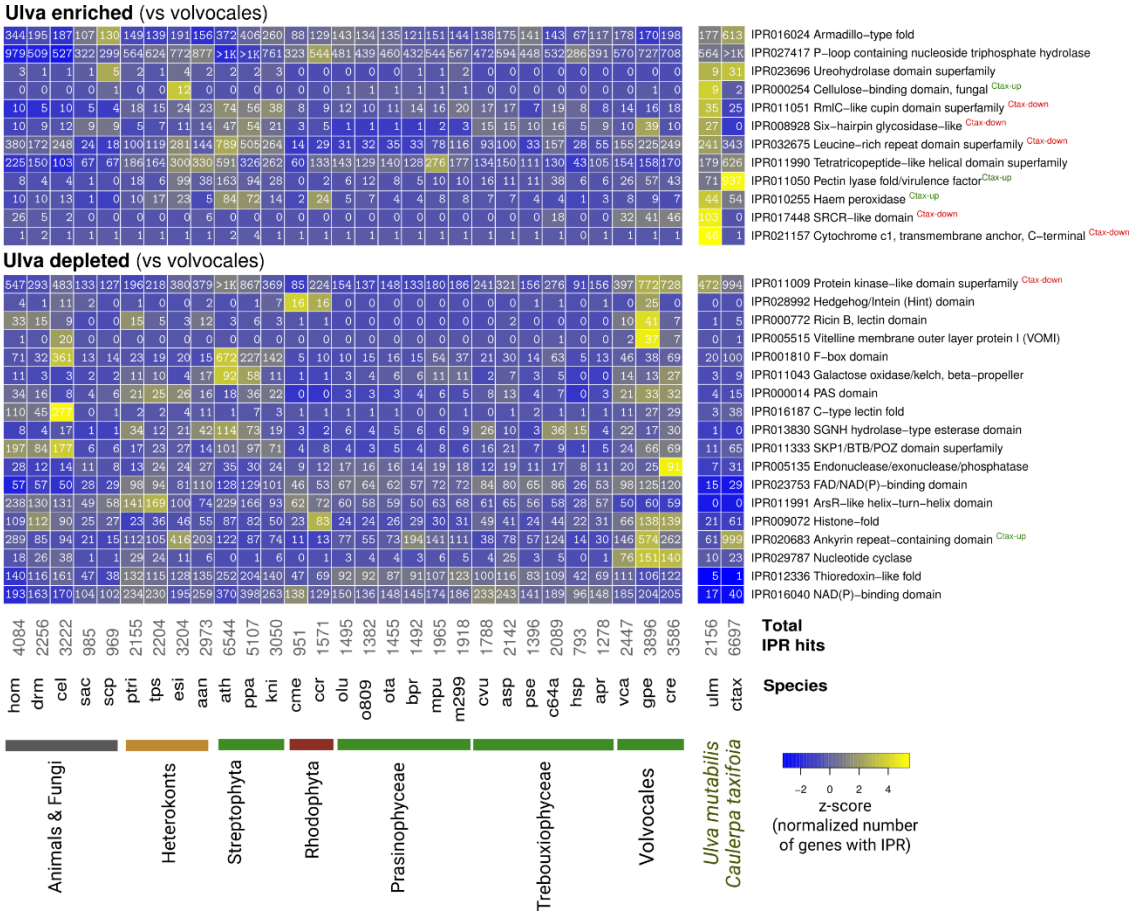
8. Fig. 4A. It would be nicer to draw a clearer color distinction between the 0 versus 1 membership category in the heat map. For example, an empty white box or light red box in cases where a family is absent will make this category stand out better.

We have adapted the coloring scheme and updated the figure. Black is absent.

9. Fig. 5. I like the idea behind this analysis but the presentation and the analysis could benefit from more careful curation. (i) Consider the repeated instances of histone related proteins in the top panel (volvocine enriched) which really represent the same result being reported multiple times with slightly different IPRs. This issue distorts the data and may mask other interesting results. A similar duplication issue exists for Ulva-enriched IPRs, e.g. glucose-methanol-choline oxidoreductaxe, Haem peroxidase, SRCR domains, integrase domain. This overcounting can be avoided by condensing IPRs into meta-categories when they are shared within the same protein. (ii) please include total number of hits and total instances for each IPR when reporting data, not just the Z-score.

We agree with this comment and reran the analyses. We took advantage of the hierarchical structure of Interpro domains (IPR) and merged those IPR domains which also turned out to be significant at the Super Family-level (e.g. Histons). By doing so we avoided incorporating redundant interpro domains. The significantly depleted and enriched IPR domains that are not grouped in higher-level Superfamilies are retained as separate entries.
Lastly, we also included the *Caulerpa* transcriptomic dataset in the analyses.

The total number of hits for each IPR is added to the figure. We did not include the number of IPR hits per species since this resulted in an overload of information which was far from esthetically pleasing. For the reference we add our attempt below. If the reviewer insist we may still add these numbers, but it is not our preference.



10. Not much was said about the pectin lyase-kinase fusion encoding genes described briefly. This seems like it could be really interesting in terms of signaling or ECM regulation. Is this a novel domain combination and how similar are the different family members for this protein type in either the kinase or pectin lyase regions? Are they the result of duplications or convergence or both?

We have adapted and expanded the paragraph on kinases and pectin lyases. In short, the coupling of kinase domains with pectin lyases is observed only in *Ectocarpus* (1 gene) and *Klebsormidium*.

11. On Pg. 12 the sentence "In Volvocales, the SRCR proteins have diverse roles in recognition of pathogen-associated molecular patterns such as bacterial interactions and sperm chemotaxis [55, 56]" implies that the function of SRCR proteins is known in Volvocales. Neither of the cited references nor any study I know about has elucidated the function(s) of SRCR proteins outside of metazoans. Please rephrase the information in this sentence.

The sentence has been rephrased. It is now clear that the reference do not refer to Volvocales but to Metazoans,

12. I found Figure S2 hard to interpret because the images for different hormones or treatments were shown at different magnification scales. If more than one magnification is required then please have a matched control sample for each so readers will be able to easily compare images. Also, the top row images are too small to see anything meaningful.

Figure S2 is replaced. All pictures were taken at the same resolution now.

Reviewer #2: Review of Ulva genome

This paper reports the genome of the multicellular green alga Ulva. Multicellularity has clearly evolved more than once in the green clade, so the genome of a clear multicellular species is important to have. One issue I have from the start is the comparison drawn to Volvox, also called by the authors a multicellular species. By no means is Volvox "multicellular" in the same way as is is Ulva. Volvox is a sphere of cells, whereas Ulva has a complete multicellular lamina, like land plants and their aquatic algal relatives (like Chara). Volvox is a collection, instead, of Chlamydomonas-like cells, and its "multicellularity" is neither homologous with Ulva, nor analogous, in my view - again, Ulva is laminar, and Volvox is a sphere of chlamydomonas-type cells. As such, it is inappropriate to say the following as regards gene family evolution: "Gain and loss, however, do not seem to be correlated with multicellularity. Volvox and Ulva, both multicellular taxa, exhibit higher loss than Chlamydomonas and Gonium, that are unicellular and colonial species, respectively.". And further, here: "10]. Apart from implying that evolution toward multicellularity progressed along different trajectories in Ulva and the volvocine algae, from a more general perspective, the absence of D-type cyclins, retinoblastome (RB) and E2F signifies that entry in the cell cycle, and the G1-S transition, are not dependent on any of these genes, similarly to yeast. As no homologues of Cln 2/3, SBF and Whi5, which mediate G1-S transition in yeast [35], are found in Ulva, we hypothesise that either a functionally analogous set of genes or an entirely different mechanism regulates Ulva S-phase entry." Instead, it may be no surprise that true multicellularity (Ulva, Chara, land plants) has different genomic basis than a small ball of chlamydomas-like cells. In other words, go ahead and make the points about the

gene family differences, but please do not claim that Volvox is in any way multicellular in the same sense as Ulva.

> While we agree in general with the reviewer's point of view, it is also important to realize that discussions on what defines multicellularity are often contentious and conditional on definitions of simple or complex multicellularity. We find this debate especially interesting in the green part of the tree of life and especially in green seaweeds where macroscopic growth and the overall complexity of an organism is in many cases partly or even completely disconnected from multicellularity (e.g. *Acetabularia*, *Caulerpa*, *Codium*, *Valonia*, …).
> Is *Volvox* multicellular? This is a fair question. The genome would prove you right, it did not reveal a lot of multicellular signatures and as matter of fact turned out to be very similar to *Chlamydomonas*. We tried to accommodate your point of view by differentiating between unicellular (*Chlamydomonas*), colonial (*Tetrabaena*, *Gonium*) and restrict the term multicellular to *Volvox*.

On the genome itself, the N50 seems reasonable at 600 Kb, and the genome size likely close to prediction. The gene annotation similarly sounds reasonable with >90% BUSCO completeness, and the additional steps taken to assess completeness are convincing enough, for example RNAseq evidence matching >90%. This said, as the authors no doubt recognize, it is difficult to know what really "should" and "should not" be in the genome with respect to other taxa. As such, when the authors note that a gene is "missing" (in their long laundry list - it is a bit tediously long), it is important to really be sure of that to make a strong statement. Please be very careful with how such conclusions are drawn. I think it would also be extremely important to BLAST against an unannotated version of the genome to see if "missing" sequences might be there yet not called as gene models for some reason. Also one can run BUSCO against unannotated genomes. Also protein domains, could run as TBLASTX against the naked genome. Please do this, as gene model prediction could be much more incomplete than realized, since nobody really knows what to expect for a green alga so distantly related to sequenced genomes. For the key genes that the authors present as (a really long laundry list…) of stories, please do everything under the sun possible to ensure that genes truly aren't there in the assembly somewhere.

> Claims of absence of specific genes or pathways have been thoroughly checked on the protein dataset and verified by blasting against the genome itself as well as the transcriptomic data (tblastx).

I also have some issues with the enrichment statistics presented by the authors - have "significant" enrichments been properly corrected for multiple tests? This is vital, as significance could disappear with a Bonferroni correction.

> The p-values already have been corrected for multiple tests. We have clarified this in the method section "Each InterPro domain was tested for enrichment or depletion in Ulva compared with the Volvocales by Fisher's exact tests with a false discovery rate correction (Benjamini-Hochberg FDR method) of 0.05."

On horizontal gene transfer, I agree that phylogenomic approaches can be very suggestive of cases of putative HGT. But there is absolutely no substitute to actually locating the proposed HGT genes in the genome - within contigs, and surrounded by clearly native DNA. In my opinion, an effort to examine native-to-HGT-to-native junctions within the Ulva assembly are necessary to convince the HGT audience in these days of long read genomes, where single reads should even pass through such junctions. Please search for representative long reads that show such cases, match them to the assembly, and then present a far stronger argument.

> We refer to our answer of a similar comment by the first reviewer. In short, we have mapped the HGT genes on the raw PacBio reads to verify if these genes are genuinely part of eukaryotic sequence data rather than contaminations. A visual representation of these analyses can be downloaded using the following link: https://bioinformatics.psb.ugent.be/gdb/ulva/HGT/mapping.zip

As I've mentioned a couple of times above, the paper reads a lot like a laundry list of presumed gains/losses of genes, and what these changes might mean. Some of the hypotheses given are, in my opinion, considerably more speculative in nature than the way they are stated in the text. Soften please, as appropriate according to editor. Also, given the long presentation, the paper could be shortened if space is at a premium by reducing the number of speculative stories.

Insights into the Evolution of Multicellularity from the Sea Lettuce Genome

Authors: Olivier De Clerck[1]*, Shu-Min Kao[2,3], Kenny A. Bogaert[1], Jonas Blomme[1,2], Fatima Foflonker[4], Michiel Kwantes[5], Emmelien Vancaester[2,3], Lisa Vanderstraeten[6], Eylem Aydogdu[2,3], Jens Boesger[5], Gianmaria Califano[5], Benedicte Charrier[7], Rachel Clewes[8], Andrea Del Cortona[1,2,3], Sofie D'Hondt[1], Noe Fernandez-Pozo[9], Claire M. Gachon[10], Marc Hanikenne[11], Linda Lattermann[5], Frederik Leliaert[1,12], Xiaojie Liu[1], Christine A. Maggs[13], Zoë A. Popper[14], John A. Raven[15,16], Michiel Van Bel[2,3], Per K.I. Wilhelmsson[9], Debashish Bhattacharya[4], Juliet C. Coates[8], Stefan A. Rensing[9], Dominique Van Der Straeten[6], Assaf Vardi[17], Lieven Sterck[2,3], Klaas Vandepoele[2,3,19], Yves Van de Peer[2,3,18,19], Thomas Wichard[5], John H. Bothwell[20]

[1] Biology Department, Ghent University, 9000 Ghent, Belgium.
[2] Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, 9052 Ghent, Belgium.
[3] VIB Center for Plant Systems Biology, Technologiepark 927, 9052 Ghent, Belgium.
[4] Department of Biochemistry and Microbiology, Rutgers University, New Brunswick, NJ 08901, USA.
[5] Institute for Inorganic and Analytical Chemistry, Jena School for Microbial Communication, Friedrich Schiller University Jena, Lessingstr. 8, 07743 Jena, Germany.
[6] Laboratory of Functional Plant Biology, Department of Biology, Ghent University, 9000 Ghent, Belgium.
[7] Morphogenesis of Macroalgae, UMR8227, CNRS-UPMC, Station Biologique, Roscoff 29680, France.
[8] School of Biosciences, University of Birmingham, Birmingham, UK.
[9] Faculty of Biology, University of Marburg, Germany.
[10] Scottish Association for Marine Science, Scottish Marine Institute, Oban, PA37 1QA, UK.
[11] InBioS-Phytosystems, University of Liège, 4000 Liège, Belgium.
[12] Botanic Garden Meise, Nieuwelaan 38, 1860 Meise, Belgium.
[13] School of Biological Sciences and Queen's University Marine Laboratory Portaferry, Queen's University Belfast, Northern Ireland, BT7 1NN, UK.
[14] Botany and Plant Science, and, Ryan Institute for Environmental, Marine and Energy Research, School of Natural Sciences, National University of Ireland Galway, Galway, Ireland.
[15] Division of Plant Sciences, University of Dundee at the James Hutton Institute, Dundee, DD2 5DA, UK.
[16] School of Biological Sciences, University of Western Australia (M048), 35 Stirling Highway, WA 6009, Australia.
[17] Department of Plant and Environmental Sciences, Weizmann Institute of Science, Rehovot 76100, Israel.
[18] Department of Genetics, Genomics Research Institute, University of Pretoria, Pretoria 0028, South Africa.
[19] Bioinformatics Institute Ghent, Ghent University, Technologiepark 927, 9052 Ghent, Belgium
[20] School of Biological and Biomedical Sciences and Durham Energy Institute, Durham University, Durham, UK.
* Correspondence: olivier.declerck@ugent.be

Summary

We report here the 98.5 Mbp haploid genome (12,924 protein coding genes) of *Ulva mutabilis*, a ubiquitous and iconic representative of the Ulvophyceae or green seaweeds. *Ulva*'s rapid and abundant growth makes it a key contributor to coastal biogeochemical cycles; its role in marine sulfur cycles is particularly important because it produces high levels of dimethylsulfoniopropionate (DMSP), the main precursor of volatile dimethyl sulfide (DMS). Rapid growth makes *Ulva* attractive biomass feedstock, but also increasingly a driver of nuisance 'green tides'. Additionally, ulvophytes are key to understanding evolution of multicellularity in the green lineage. Furthermore, morphogenesis is dependent on bacterial signals, making it an important species to study cross-kingdom communication. Our sequenced genome informs these aspects of ulvophyte cell biology, physiology and ecology. Gene family expansions associated with multicellularity are distinct from those of freshwater algae. Candidate genes are present for the transport and metabolism of DMSP, including some that arose following horizontal gene transfer from chromalveolates. The *Ulva* genome offers, therefore, new opportunities to understand coastal and marine ecosystems, and the fundamental evolution of the green lineage.

Key words: green seaweeds, multicellularity, phytohormones, DMSP, DMS, *Ulva*

Introduction

Transitions from microscopic, unicellular life forms to complex multicellular organisms are relatively rare events but have occurred in all major lineages of eukaryotes, including animals, fungi, and plants [1-3]. Algae, having acquired complex multicellularity several times independently, provided unique insights into the underlying mechanisms that facilitate such transitions. In the green lineage, there were several independent transitions to multicellularity. Within Streptophyta, the origin of complex land plants from a green algal ancestor was preceded by a series of morphological, cytological, and physiological innovations that occurred long before the colonization of land [4, 5]. Within the Chlorophyta, the transition from uni- to multicellularity has occurred in several clades and has been studied extensively in the volvocine lineage [6, 7]. Comparative genomic analyses between unicellular (*Chlamydomonas*), colonial (*Gonium*, *Tetrabaena*) and multicellular species (*Volvox*) revealed protein-coding regions to be very similar [8-10], with the notable exceptions of the expansion of gene families involved in extracellular matrix (ECM) formation and cell-cycle regulation.

The Ulvophyceae or green seaweeds represent an independent acquisition of a macroscopic plant-like vegetative body known as a thallus. Ulvophyceae display an astounding morphological and cytological diversity [11], which includes unicells, filaments, sheet-like thalli, as well as giant-celled coenocytic or siphonous seaweeds [12, 13]. *Ulva*, or sea lettuce, is by far the best-known representative of the Ulvophyceae. *Ulva* is a model organism for studying morphogenesis in green seaweeds [14]. The *Ulva* thallus is relatively simple, with small uninucleate cells and a limited number of cell types. *Ulva* exists in the wild in two forms, i.e. as flattened blades that are two cells thick or tubes one cell thick (Fig. 1A). Both forms co-occur in most clades as well as within single species [15-17]. These morphologies, however, are only established in the presence of appropriate bacterial communities [18]. In axenic culture conditions, *Ulva* grows as a loose aggregate of cells with malformed cell walls (Fig. 1B). Only when exposed to certain bacterial strains (e.g., *Roseovarius* and *Maribacter*), or grown in conditioned medium, is complete morphogenesis observed [18, 19]. Thallusin, a chemical cue inducing morphogenesis, has been characterized for the related genus *Monostroma* [20] and additional substances that induce cell division (*Roseovarius*-factor) and cell differentiation (*Maribacter*-factor) have been partially purified from *Ulva*-associated bacteria [18]. The value of *Ulva* as a model organism for green seaweed morphogenesis is enhanced by its tractability for genetic analyses. Under laboratory conditions, individuals readily complete the life cycle, which involves an alternation of morphologically identical haploid gametophytes and diploid

sporophytes [14]. Furthermore, a stable polyethylene glycol (PEG)-based genetic transformation system is available [21].

Ulva genus is widely distributed along tropical and temperate coasts, and several species penetrate into freshwater streams and lakes. Under high nutrient conditions, *Ulva* can give rise to spectacular blooms, known as “green tides”, often covering several hundreds of kilometres of coastal waters. Beached algae may amount to a million tons of biomass and smother entire coastlines and negatively impact tourism and local economies [22, 23]. Although not toxic, green tides have led to fatalities due to the hydrogen sulphide that is formed when they decay. Despite the harmful consequences of *Ulva*’s rapid growth rate, beneficial aspects include exploitation of its biomass; e.g., for biofuel production [24, 25], protein for animal feed [26], and the removal of excess nutrients in integrated multitrophic aquaculture systems [27-29].

Genomic resources of Ulvophyceae that could inform on this independent transition to multicellularity and the evolution of the cyto-morphological diversity are limited to a transcriptomic study of *Ulva linza* [30] and a description of the mating type locus of *U. partita* [31]. In addition, Ranjan et al. [32] studied the distribution of transcripts in the thallus of the siphonal green seaweed *Caulerpa* .Here, we present the first whole genome sequence of an *Ulva* species, *U. mutabilis* Føyn (Fig. 1). This species is phylogenetically closely related to *U. compressa* Linnaeus (Fig. S1), a widespread species known to form nuisance blooms. The sequenced strain was isolated from southern Portugal [33] and has been successfully maintained in culture. Several important aspects of its biology, including cell cycle, cytology, life cycle transition, induction of spore and gamete formation and bacterial-controlled morphogenesis, have been studied in detail over the past 60 years ([reviewed in 14, 17]). The *Ulva* genome, in combination with the availability of a genetic toolkit and developmental and life cycle mutants, will spur new developmental research in green seaweeds which is important for the success of these seaweeds as an aquaculture resource as well as the possible control of blooms. Here, our analyses of the *Ulva* genome reveal features that underpin the development of a multicellular thallus, and expansion of gene families linked to the perception of photoperiodic signals and abiotic stress, which are key factors for survival in intertidal habitats. Furthermore, we unveil key genes involved in the biosynthesis of dimethylsulphoniopropionate (DMSP) and dimethyl sulfide (DMS), important signaling molecules with a critical role in the global sulfur cycle.

Results and Discussion

Genome sequencing and gene family evolution

The genome size of *U. mutabilis* was estimated by flow cytometry and k-mer spectral analysis to be around 100 Mbp. In total, 6.9 Gbp of PacBio long reads were assembled into 318 scaffolds (98.5 Mbp), covering 98.5 % of the estimated nuclear haploid genome (Table 1, Table S1). To increase the accuracy of the genome sequence at single-base resolution, the scaffolds were polished using PacBio and Illumina paired-end reads. We predicted 12,924 protein-coding genes, of which 91.8 % were supported by RNA-Seq data. Analyses of genome completeness indicated that the genome assembly captures at least 92% of eukaryotic BUSCO [34] and a 0.968 completeness score by pico-PLAZA Core Gene Family [35] of the protein coding genes (Table S2). Annotation of repetitive elements resulted in 35% of the genome being masked. Among the identified repeats, 74% were classified as known or reported repeat families, with long-terminal repeats (LTRs) and long interspersed elements (LINEs) being predominant, representing 15.3 Mbp and 9.3 Mbp, respectively (Table S3).

Table 1 | Summary statistics for the *Ulva mutabilis* genome.

Genome size (Mbp)	**98.5**
Scaffold N50 (Mbp)	0.6
% G and C	57.2
Number of protein coding genes	12,924
Gene density (genes/Mb)	131.2
Average intron per gene	3.9
Average exon length (bp)	303.1
Average intron length (bp)	368.6
% of genes with introns	85.0

The *Ulva* genome size is intermediate between sequenced genomes of Chlorophyceae and Trebouxiophyceae (Fig. 2). The number of predicted genes and gene families are markedly lower compared to most Chlorophyceae, including the volvocine algae (*Chlamydomonas*, *Gonium, Tetrabaena* and *Volvox*), but higher than the Trebouxiophyceae and prasinophytes (Fig. 2, 3). The relative gene family sizes, however, are roughly equal between *Ulva* and volvocine algae (Fig. S2).

. A phylogenetic tree inferred from a concatenated alignment of 58 nuclear protein coding genes (totaling 42,401 amino acids) supports a sister-group relationship of *Ulva* with the Chlorophyceae in the crown chlorophytes (Fig. 3). This topology corroborates earlier phylogenetic hypotheses based on multigene organelle datasets (reviewed in [11]). The divergence of *Ulva* and the *Chlamydomonas* – *Gonium* – *Volvox* clade (Chlorophyceae, Volvocales) from their common ancestor coincided with substantial gain and loss of gene families in both lineages. Gain and loss, however, do not seem to be correlated with multicellularity. Multicellular taxa exhibit higher loss than unicellular ones.

Evolution of multicellularity

Ulva develops from gametes or zoospores into a multicellular thallus consisting of three main cell types (rhizoid, stem and blade cells). After a first division, the basal cell gives rise to a rhizoidal cell and the apical cell [36]. Morphogenesis is governed by a progressive change in cell cycles. The growth rate of the basal cells decreases after few cell cycles resulting in a small holdfast, whereas division of blade cells becomes synchronised by a light:dark cycle [36, 37]. The formation of a complex multicellular thallus in *Ulva* is therefore largely a question of how cell size and division are controlled and many morphological mutants of *Ulva mutabilis*, including slender, apparently underwent changes in cell cycle regulation [36].

Evolution of a complex morphological thallus is often linked to expansions in genes that are required for cell signaling, transcriptional regulation and cell adhesion, [3, 38, 39]. The *Ulva* genome encodes 251 proteins involved in transcriptional regulation, a comparatively low number for a green alga, which is also reflected in a low fraction of such proteins encoded by the genome (1.94% when compared to the average of 2.66% in green algae). *Ulva* lacks ten families of transcription factors (TFs) and two families of transcriptional regulators (TR) that are present in other green algae (Suppl.Data1). Furthermore, the existing transcription-associated protein families are, on average, smaller than those in other green algae (Fig. 4A).

Among the most remarkable gene families that have been lost are genes of the retinoblastoma/E2F pathway and associated D-type cyclins. Comparative genomics of volvocine algae revealed that the co-option of the retinoblastoma cell cycle pathway is a key step towards multicellularity in this group of green algae [8-10]. Apart from implying that evolution toward multicellularity progressed along different trajectories in *Ulva* and the volvocine algae, from a more general perspective, the absence of D-type cyclins, retinoblastoma (RB) and E2F signifies that entry in the cell cycle, and the G1-S transition, are not

dependent on any of these genes, similarly to yeast. As no homologues of Cln 2/3, SBF and Whi5, which mediate G1-S transition in yeast [40], are found in *Ulva*, we hypothesise that either a functionally analogous set of genes or an entirely different mechanism regulates *Ulva* S-phase entry. Both retinoblastoma (RB) and E2F were found in the transcriptome of the siphonous ulvophyte *Caulerpa* [32], making it unclear at present to which extent the absence of the retinoblastoma/E2F pathway is a feature unique to the *Ulva* lineage. Other aspects of the cell cycle are more in line with other green algae (Suppl.Data1), be it that the single CDKA homolog (UM001_0289) contains a modified cyclin-binding motif, PSTALRE, instead of the evolutionarily conserved PSTAIRE motif. While not uncommon in eukaryotes *Ulva* is the first member of the green lineage with variation in the PSTAIRE motif [40].

Contrary to the expectation for a multicellular organism [39, 41], few TF families are expanded in *Ulva* (Fig. 4A). A notable exception is CONSTANS-LIKE transcription factors (CO-like) of which *Ulva* has five genes, whereas all other analysed algae encode between zero and two (Fig. 4A,B). These CONSTANS-LIKE transcription factors are characterised by one or two (Group II or III) Zinc finger B-boxes and a CCT protein domain. Both protein domains are involved in protein-protein interactions and the CCT domain mediates DNA-binding in a complex with HEME ACTIVATOR PROTEIN (HAP)-type transcription factors in *Arabidopsis* [39, 42-44]. *Ulva* CO-like proteins form a single clade within other algal lineages (Fig. 4B). In addition to the five CO-like transcription factors, functionally related proteins containing only a B-Box domain, similar to Group V Zinc Fingers B-Box [43, 45] or only a CCT domain, belonging to the CCT MOTIF FAMILY [46], are also expanded in *Ulva*. B-box Zinc Fingers and CMF proteins in angiosperms have been implicated in developmental processes such as photoperiodic flowering [47], regulation of circadian rhythm [48], light signalling [49] and abiotic stress [50]. The control of light and photoperiod signalling is conserved in the moss *Physcomitrella patens* [51], and the green algae *Chlamydomonas reinhardtii* and *Ostreococcus tauri* [52, 53]. Moreover, the CO-like TFs are one of the families potentially involved in the establishment of complex multicellularity in green algae and land plants [54]. Genome-wide mapping of *Ulva* CO-like genes and functionally related genes indicates that the majority (60%) originated through tandem duplication in *Ulva* (Fig. 4C). Although the functions of these proteins will need to be confirmed experimentally, the *CO*-like and *CMF* genes in *Ulva* could be involved in the integration of a multitude of environmental signals in a highly dynamic intertidal environment.

A total of 441 protein kinases was identified in the *Ulva* genome, representing ca. 3-4% of all protein coding genes, a smaller proportion than in *Arabidopsis* (4% [55]), but more than in *Chlamydomonas* (2.3% [56, 57]) and the brown seaweed *Ectocarpus* (1.6% [58]). The largest subfamily of *Ulva* kinases has

similarity to PKnB kinase (Suppl. Data1), a “eukaryotic-like” serine/threonine kinase originally discovered in bacteria. Around 20 of the PKnB kinases possess a clear transmembrane (TM) domain and an extracellular/adhesion domain: either Kringle (IPR000001), Fasciclin (IPR000782), Pectin lyase (IPR012334, IPR011050), and so represent good candidates for *Ulva* receptor-kinases with a potential role in environmental sensing and/or developmental signalling. The pectin lyase domain for example may, through activity of the protein on *Ulva* cell wall components, contribute to desiccation resistance [59] and the growth and development of a multicellular thallus [60, 61]. Although the highlighted extracellular domains are present throughout green algae, including the green seaweed *Caulerpa* (Fig. 5), the coupling of these domains with a kinase domain is a feature observed in a few species only. The Kringle-kinase domain combination was initially discovered in animals, in receptor tyrosine kinase-like orphan receptors (RORs), which use Wnt signalling proteins as ligands and function in multicellular development, neuronal outgrowth, cell migration and- polarity [62]. Our analysis additionally finds this domain combination in *Ulva* and unicellular prasinophyte green algae (*Ostreococcus*). The Fasciclin-kinase combination is unique to *Ulva*, while the pectin-lyase-kinase combination is found only in the multicellular algae *Ulva*, *Klebsormidium* and *Ectocarpus.* It is possible that *Ulva* Kringle-TM-kinase gene families arose via divergent evolution from a common ancestor, based on sequence similarity of family members and the close proximity of some family members on single DNA scaffolds. The pectin lyase-TM-kinase family proteins are more divergent in sequence and structure (including kinase-TM-pectin lyase proteins) and are more likely to have arisen by dynamic gene fusions or conversions.

In addition to extracellular protein domains associated with intracellular kinases, *Ulva* also shows a significantly enriched diversity in protein domains associated with the extracellular matrix (ECM) and cell surface, relative to its sequenced sister taxa. Other enriched domains associated with the ECM are Scavenger Receptor and Cysteine Rich domain (SRCR) proteins (IPR001190, IPR017448), which are absent from land plants but present in animals and Volvocales [57]. In Metazoans, the SRCR proteins have diverse roles in recognition of pathogen-associated molecular patterns such as bacterial interactions and sperm chemotaxis [63, 64]. SRCR proteins are likely to have an early evolutionary role in cell-cell recognition or aggregation [65]. The germin domain (IPR000129) is among the abundant gained domains and is also encountered in streptophytes, where it is linked to regulating cell wall properties including extensibility [66] and defense [67]. Germins, glycoproteins that occur ubiquitously in land plants, typically contain a RmlC-Like cupin domain fold (IPR011051) which is also significantly enriched in comparison to Volvocales.

The gametolysin/MEROPS peptidase family M11 of VMPs (*Volvox* matrix metalloproteases) (IPR008752), similar to mammalian collagenases and crucial to ECM remodeling (Hallman et al. 2001), is enriched in *Volvox* relative to *Chlamydomonas* [9]. While the M11 class of peptidases is absent in *Ulva*, the related MEROPS M8/leishmanolysin peptidase domain (IPR001577) is 23-fold enriched compared to *Chlamydomonas.* Also, 28 collagen triple helix repeat (IPR008160) proteins with a predicted extracellular location have been detected with a G-N/D-E pattern, suggesting the presence of collagen-like innovations in the ECM of *Ulva*.

Phytohormones

Growth and development in land plants are modulated by plant hormones. Morphogenesis of *Ulva* into a blade- or tube-like thallus may therefore also require specific plant hormones as suggested by previous experimental studies [68]. However, because bacteria can produce plant hormones, it has not previously been clarified whether hormones detected in *Ulva* arise from the alga or its associated bacteria. Furthermore, because the *Roseovarius*-factor resembles a cytokinin and the *Maribacter*-factor acts similar to auxin, bacteria-derived compounds may contribute to the development of the multicellular *Ulva* thallus [18]. We, therefore, investigated if biosynthesis pathways of phytohormones are present in *Ulva* and/or the associated bacteria and tested whether hormones can replace the bacterial morphogenetic factors. However, none of the identified hormones triggers growth or development of gametes and young propagules in standardized bioassays (Fig. S3). Homologs of plant hormone biosynthesis genes provided strong evidence for biosynthesis of abscisic acid (ABA), ethylene (ET) and salicylic acid (SA) in *Ulva* (Fig. 6, Suppl. Data1). Corroborating these results, we found that both xenic and axenic *Ulva* produce ABA, ET and SA, but also auxin (IAA) and gibberellin (GA_3) (Table S4). Measurements of IAA in axenic cultures are more difficult to reconcile with the *Ulva* gene content, as little or no evidence was found for indole-3-pyruvic acid (IPA), tryptamine (TAM), and indole-3-acetamide (IAM) pathways. Biosynthesis of IAA seems most likely to involve the conversion of indole-3-acetaldoxime and indole-3-acetaldehyde by AMI1 and AAO1, respectively. The presence of GA_3 remains equivocal. In line with earlier results of Gupta et al. [69], traces of GA_3 could be detected in both axenic and xenic slender cultures, but not in wildtype. At present, it is unclear how traces of GA_3 can be reconciled with the gene content of *Ulva* since enzymes mediating ent-kaurene biosynthesis in the plastid (CPP SYNTHASE, ENT-KAURENE SYNTHASE) are missing from the genome. We found that the associated bacteria *Roseovarius* sp. MS2 and *Maribacter* sp. MS6 produce, IAA, GA_3, ET, SA and the cytokinins (iP) (Table S4). In contrast to the evidence of biosynthesis of several

plant hormones, the corresponding receptors, as characterized in angiosperms, were absent, corroborating the comparative genomic analyses by Wang et al. [70]. These authors found little evidence for the emergence of homologous plant hormone signaling pathways outside the charophyte lineage. These findings, however, do not preclude hormonal or other functions by different pathways and interdependencies as demonstrated for diatoms [71] and brown algae [72].

Macroalgal-bacterial interactions

Ulva relies on interactions with bacteria for both settlement of zoospores [73, 74] and morphogenesis of the thallus [14, 18]. This close association with bacteria, however, does not seem to have resulted in significantly higher levels of horizontal gene transfer (HGT). Using a phylogenomic pipeline to investigate the extent of HGT in the *Ulva* genome, we found 13 well-supported cases of HGT of prokaryotic origin (Table S5). Although this number is not exceptionally high, it is remarkable that detected HGT events are in more than half of the cases followed by gene family expansion. The most striking case is present by Haem peroxidases , belonging to the peroxidase-cyclooxygenase superfamily, of which *Ulva mutabilis* has 36 copies. The expansion results from a single HGT event. Peroxidases, involved in scavenging H_2O_2, are part of *Ulva*'s antioxidant machinery to cope with environmental challenges common to intertidal habitats such as excessive light, hypersalinity and dehydration [30, 75], but may also have important functions in cell wall modifications [76]. In plants peroxidases have been demonstrated to cross-link extensins resulting in a stiffening the primary cell wall. In this process pectins act as an anchor for the peroxidases. Noteworthy, genes coding for extensins and pectin-like polysaccharides are very prominent in the *Ulva* genome (Suppl.Data1). Predictions of localization corroborate an extracellular function of 21 of these peroxidases, while 5 are targeted to the mitochondrion.

One such interaction may involve biosynthesis of siderophores and iron uptake. Pilot studies have recently revealed that bacteria tightly associated with *U. mutabilis* release unknown organic ligands which complex iron [78]. Released microbial bacterial siderophores become, therefore, part of the organic matter in the chemosphere and contribute to the recruitment of iron within the tripartite community in *U. mutabilis*. It was hypothesized that *Ulva* uses siderophores as public goods within a bacterial-algal mutualism, where heterotrophic bacteria are fed by the algae through the release of carbon sources. Genome mining for iron uptake genes suggests that *Ulva* can acquire iron, maintained in solution and bioavailable by bacterial

siderophores, via reduction by a ferric chelate reductase, re-oxidation by a multicopper ferroxidase and finally uptake by a transferrin-like protein (Suppl.Data1).

Dimethylsulfoniopropionate

Dimethylsulfoniopropionate (DMSP) has been identified as a major metabolite in *Ulva*. In addition to its role as an osmolyte and cryoprotectant, DMSP also appears to play a direct role in cross-kingdom signalling between *Ulva* and associated bacteria [79]. *Roseovarius* sp. strain MS2 is chemotactically attracted to DMSP . Bacteria use the algal DMSP signal as a mechanism for detecting the presence of a photoautotrophic organism releasing various carbon sources such as glycerol [79]. In turn, *Roseovarius* sp. promotes the growth of *Ulva* by a cytokinine-like substance. DMSP works thus as an important chemical mediator for macroalgal-bacteria interactions [79]. Comparative genome analysis reveals genes with a putative function in biosynthesis, catabolism and transport of DMSP.

Biosynthesis of dimethylsulfoniopropionate (DMSP) in algae (incl. *Ulva*) was demonstrated to follow a route entirely different from that in higher plants [80]. In a four-step pathway methionine is first transaminated, reduced, and methylated to give rise to 4-methylthio-2-oxobutyrate (MTOB), 4-methylthio-2-hydroxybutyrate (MTHB) and 4-dimethylsulfonio-2-hydroxy-butyrate (DMSHB), respectively (Fig. 7A). Finally, oxidative decarboxylation of DMSHB yields DMSP. The methylation of MTHB to DMSHB is known only in association with DMSP synthesis and is considered to be the committing step in the pathway [81]. The *DSYB* gene, a eukaryotic homologue of bacterial *dsyB* [82], has recently been shown to mediate the conversion of MTHB to DMSHB in several eukaryotic DMSP-producing algae [83]. A *DSYB* homologue could not be found however in the *Ulva* genome. The gene was also absent in the genomes of other DMSP-producing diatoms *Phaeodactylum tricornutum* and *Thalassiosira pseudonana [82]*, indicating that multiple biosynthesis pathways exist in eukaryotes. Here we identify an *Ulva* S-adenosyl-L-methionine-dependent methyltransferase (UM036_0102), homologous to a candidate methyltransferase in *Fragilariopsis cylindrus* (CCMP 1102, accession 212856a) involved in DMSP production [84]. Homologues of UM036_0102 are notably absent from non-DMSP-producing Chlorophyceae and Trebouxiophyceae. Expression analysis reveals that UM036_0102 is upregulated 7.0-fold (± 1.2 SD) at low temperatures (8°C) (Fig. 7B). Expression values correlate with DMSP concentrations which are significantly higher at low temperatures (5.75 ± 1.8 mg / g FW) compared to 18°C (3.14 ± 1.05 mg / g FW; Student's t-Test $p < 0.001$, $n = 12$), conform the potential role of DMSP as a cryoprotectant. Expression of UM052_056, a SAM methyltransferase with homologues in volvocine algae which was included as a control, was not increased. The combination of comparative analyses and expression data

make UM036_0102 a credible candidate for the protein mediating the methylation step in the biosynthesis of DMSP, alternative to *DSYB* in eukaryotes.

Relevant in the context of DMSP synthesis are 4 putative BCCT-type family symporters/antiporters (IPR000060) which have been demonstrated to be involved in the import of DMSP in several bacteria [85, 86]. The presence of DMSP transporters has been postulated in diatoms [87, 88] and *Ulva* [89]. In *Ulva,* which shifts from DMSP synthesis towards DMSP uptake under S deficiency, a putative DMSP transporter has been shown to be a Na^+/DMSP symporter, which points towards BCCT transporters [89]. Homologues of the putative *Ulva* BCCT transporters are also present in diatoms, some marine prasinophytes and opisthokonts, but are notably absent from the fresh water Trebouxiophyceae and Chlorophyceae. Two out of four BCCT transporters are significantly upregulated at low temperatures UM033_147 and 150), while one is significantly down-regulated (UM033_146) (Fig. 7B).

The *Ulva* genome encodes two copies of a DMSP lyase, the enzyme responsible for forming DMS from DMSP (Fig. 7A), which was originally identified in *Emiliania huxleyi* [90]. *Ulva* may thus regulate the chemoattraction of bacteria by the *de novo* synthesis of DMSP and its decomposition to DMS through the intrinsic DMSP lyase. The production of DMSP and DMS by *Ulva* [80], and its ecological implications for the natural sulfur cycle [91] and climate regulation (e.g. cloud condensation) [92], are an important topic of research. The involvement of a DMSP lyase in the production of DMS has been demonstrated in the past [93, 94], but the enzymes remained unidentified so far. Furthermore, the production of DMS by *Ulva* rather than by associated bacteria was unequivocally demonstrated only recently [79]. Phylogenetic analyses resolved both *U. mutabilis* DMSP lyases in a clade with several haptophytes (including Alma3, 6 and 7 of *Emiliana huxleyi*) and dinoflagellates but also the scleractinian coral *Acropora* (Fig. 7C). The most likely explanation for the presence of the Alma DMSP lyases in *Ulva*, the first in any green algal genome, is thus a HGT from the "chromalveolate" lineage rather than from bacteria. UM030_0039 was 13-fold downregulated at low temperatures (Fig. 7B). The expression of UM021_0036 did not vary significantly across treatments, which may indicate a function disconnected from DMS formation.

Conclusion

Genome analysis of *Ulva mutabilis* reveals several key insights into the biology of green seaweeds. Most importantly, we demonstrate that a long independent evolution from *Ulva*'s evolution from a unicellular and freshwater ancestral chlorophyte, has resulted in marked changes in the genetic basis underlying the

most important aspects of the species' biology (e.g. cell cycle control). Although the unicellular–multicellular transition in volvocine algae relies on completely different proximate causes such as co-option of the retinoblastoma cell cycle pathway, interesting parallels can be drawn concerning expansion of ECM-related gene families and their putative role in environmental signaling. Additional advances in the precision and efficiency of reverse genetics and genome-editing techniques, combined with the intrinsic characteristics enabling *Ulva* to grow easily under laboratory conditions as clonal and haploid plants, offer exciting prospects to make *Ulva* the model organism for green seaweeds, complementary to *Chlamydomonas*. *Ulva* is increasingly used as a crop in seaweed aquaculture. Sustainable and biosecure exploitation and domestication of *Ulva* can benefit from identification of ecotypic genetic variation underlying bloom-formation. For example, the comparison between bloom- and non-bloom-forming *Ulva* species may assist understanding of the molecular mechanisms underpinning growth and reproduction in response to environmental conditions. Furthermore, expanding comparative genomics to giant-celled green seaweeds (e.g. *Acetabularia*, *Caulerpa*, and *Cladophora*) has the potential to shed light on these curious macroscopic organismal architectures that do not rely on multicellularity for morphological patterning [95]. Lastly, the reliance of *Ulva* on bacterial cues for growth and morphogenesis makes the species an exciting model to study the evolution of cross-kingdom signaling in the marine environment.

STAR Methods

Key Resource Table => see separate file.

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Olivier De Clerck (Olivier.declerck@ugent.be).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The *U. mutabilis* strain sequenced, a wildtype gametophyte of mating type minus (wt-G(mt$^-$) (gametophyte); ([mt$^-$]; G/PS- swi$^+$; mut-; RS140$^+$; RS180$^+$)) was initially isolated from Ria Formosa, southern Portugal, by B. Føyn [33, 96]. An additional haploid strain, *slender*, a spontaneous mutant derived from the original collection [114], was selected to complement the available transcriptomes because of its fast-growing nature and the ease with which thalli can be induced for gamete formation. Strains of *U. mutabilis* are primarily maintained at the Friedrich-Schiller-Universität Jena. Gametophytes were raised parthenogenetically from unmated gametes in *Ulva* Culture Medium (UCM without antibiotics) in the presence of bacterial symbionts *Roseovarius* sp. MS2 (Genbank: EU359909) and *Maribacter* sp. MS6 (Genbank: EU359911) to secure normal thallus morphogenesis [115]. Bacterial strains are cultivated in marine broth medium (Roth, Germany). Algae were cultured at 20°C, under long day light conditions (L/D 17:7) consisting of 60–120 $\mu mol \cdot m^{-2} \cdot s^{-1}$ (50% GroLux, 50% day-light fluorescent tubes; OSRAM, München, Germany), without aeration.

METHOD DETAILS

DNA and RNA extraction and library construction

DNA for sequencing was prepared from axenic gametes. For mate pair libraries adult, non-axenic *Ulva* tissue was used. DNA extraction was performed using a CTAB protocol. The PacBio library was prepared using P6-C4.0 chemistry with size selection using a 0.75% cassette on a Blue Pippin instrument with a lower cutoff of 8 kb based on the Pacific Biosciences 20kb template protocol (ref 100-286-000-08). Libraries for short reads and mate pairs were constructed using the TruSeq DNA PCR-Free Library Preparation Kit and the Nextera Mate Pair Sample Preparation Kit, respectively (Illumina, San Diego, CA).

Total RNA was isolated from vegetative adult tissue, tissue in the process of gamete formation as well as gametes using the RNeasy Mini Kit (Qiagen). cDNA libraries were constructed using the ScriptSeq v2 RNA-Seq Library Preparation Kit.

Genome sequencing and assembly

Sequencing. Genomic DNA was sheared to produce fragments of 350 to 550 bp and sequenced on a MiSeq2000 (2x250 bp PE reads). Reads were trimmed and quality filtered using Cutadapt (v 1.2.1) and Sickle (v 1.200) [4]. Mate pair libraries were sequenced on an Illumina HiSeq 2500 (2x125 bp reads). The PacBio libraries were sequenced on a PacBio RSII instrument (five SMRTCells P6-C4 chemistry). RNA libraries were sequenced on one lane of Illumina HiSeq 2500 (2x125 bp PE reads) and one run of NextSeq 550 (1x150 bp SE reads). Table S6 presents an overview of the sequenced libraries.

Estimation of genome size. Both k-mer analysis and flow cytometry experiments were used to gauge the genome size of *Ulva mutabilis* (wild-type). Based on the concatenated Illumina paired-end libraries the genome size was estimated 93.6 Mbp by SGA PreQC [6] with default k-mer size 31 and 104.5 Mbp by the estimating process in ALLPATHS-LG [7] with k-mer size 25. For flow cytometry estimates, *Ulva mutabilis* nuclei were stained together with nuclei of the standard (*Arabidopsis thaliana*) and relative fluorescence was used to calculate the genome size as described by Hare & Johnston [8]. Fluorescence emission was collected using the S3e™ Cell Sorter (BIO-RAD). The flow cytometry measurements showed an estimated genome size of 100.2 ± 3.6 Mbp (mean ± standard error of four measurements).

Genome assembly. PacBio reads (6.9 Gbp) were assembled using Canu [9] resulting in a 98.4 Mbp assembly in 1,119 contigs. The 30x of the longest corrected reads from the Canu pipeline (N50 of 9.0 kbp) and the Illumina paired-end reads were used by MaSuRCA [10] to generate a hybrid assembly (with USE_LINKING_MATES turning on), resulting in an assembly of 108.1 Mbp, with scaffold N50 of 264.9 kbp and longest scaffold in 2.7 Mbp. A graph-based scaffolder MeDuSa [11] was used to scaffold the Canu contigs based on the MaSuRCA assembly, followed by SSPACE [12] scaffolding using all the mate-pair libraries. The super-scaffolds were first polished by PacBio reads using Arrow v2.2.1 (from SMRT Link v5.0.1) after mapping all the long reads by pbalign v0.3.1 (from SMRT Link v5.0.1). The PacBio-polished scaffolds were further improved using the paired-end Illumina reads with Pilon [13]. To eliminate putative bacterial contamination super-scaffolds were searched against the NCBI nucleotide (nt) database using BLAST megablast [14] with an (e-value < 1e-65), and assigned to taxonomy group based on the 'bestsum' rule in Blobtools [15].

De novo repeat finding and repeat masking

A *de novo* repeat identification was performed with RepeatModeler (1.0.8) [29]. Unknown elements were screened with BlastX (e-value < 1e-5) against UniRef90 database [30] (subset Viridiplantae) and removed from the repeat library if necessary. The filtered *Ulva* repeat library was used by RepeatMasker (4.0.7) [31] to mask the repetitive elements in the assembly, which resulted in 34.7 Mbp (35.28%) of the genome masked.

Gene prediction

We applied EVidenceModeler [32] to predict gene models. The consensus gene models were reconciled using the models from *ab initio* and orthology-aided predictions, transcripts reconstructed from RNA-Seq, and homologous models derived from the protein alignments of the available public resource (Fig. S4A). We used BRAKER1 v1.9 [34] to predict the gene models incorporating the RNA -Seq mapping results generated using HISAT2 v2.0.5 [33]. We further used Augustus v3.2.3 with the trained data from BRAKER1 and the protein profile extension to re-predict the gene models [35]. Protein profiles were generated by processing the missing family identified after gene family assignment using OrthoFinder v2.1.2 [36] with the following reference sequence from Phytozome v12.1: *Chlamydomonas reinhardtii* v5.5 [37], *Coccomyxa subellipsoidea* C-169 v2.0 [38], *Dunaliella salina* v1.0 [39], and *Volvox carteri* v2.1 [40]. The reference sequence of *Gonium pectorale* [41] was from NCBI (assembly ASM158458v1). In addition to the *ab initio* prediction, the RNA-Seq data were also used to reconstruct the transcripts, which consisted of consensus transcripts predicted by Scallop v0.10.2 [43] and predicted coding regions of Trinity v2.4.0 [44] assemblies (both *de novo* and genome-guided) using PASA [32]. Spliced alignments of proteins from UniRef90 (with taxonomy ID 33090) and the FrameDP-corrected [47] and predicted proteins sequence of green algal transcriptomes in the oneKP project [48] were generated using Exonerate [49] seeded by Diamond [50]. We combined the aforementioned gene models with the alignments of proteins, annotation of repetitive elements to produce a consensus gene set using EVidenceModeler. Non-coding RNA and tRNA were identified using infernal v1.1rc [51] and tRNAscan-SE v1.31 [26], respectively. All gene models were functionally annotated using InterProScan 5.27-66 [52] and uploaded to the ORCAE platform [53], enabling members of the consortium to curate and manually annotate.

Genome completeness

Completeness of the predicted *Ulva* gene space was evaluated using BUSCO [54] and the coreGF analysis [55]. Core gene families were defined as gene families shared among all *Chlorophyta* present in pico-

PLAZA v. 2.0 [56]. In total 1,815 gene families were compared by sequence similarity to the *Ulva* gene models and the reference protein sequences. We assigned each *Ulva* gene model to a particular gene family based on the top 5 hits (E-value < 1e-5). Finally, a GF score was calculated as the sum of each core family identified, counted with a weight equal to one divided by the average family size. Hence, a GF score of 1 indicates that all the core gene families were identified, while a GF score of 0 indicates that no core gene families were found. The likelihood of the presence of each core gene family was calculated for *Ulva* gene models.

Comparative genomic analyses

For the comparative genomic analyses a custom version of Pico-Plaza [103] was built containing genomes and annotations of 32 eukaryotic species (Table S7). Following an 'all-versus-all' BLASTP [116] protein sequence similarity search, both TribeMCL v10-201 [104] and OrthoMCL v2.0 [105] were used to delineate gene families and subfamilies. Collinear regions (regions with conserved gene content and order) were detected using i-ADHoRe 3.0 [106] with the following settings: gf_type = TribeMCL gene families, alignment method = gg2, number of anchor point s=5, gap size = 30, cluster gap = 35, tandem gap = 30, q-value = 0.85, probability cut-off = 0.01, multiple hypothesis correction = FDR and level_2_only = false. The phylogenetic profile of TribeMCL gene families (excluding orphans) retrieved from pico-PLAZA and the inferred species tree topology were provided to reconstruct the most parsimonious gain and loss scenario for every gene family using the Dollop program from PHYLIP v3.69 [107]. Gene family losses and gains were further analysed by examining the associated Gene Ontology (GO) terms and InterPro domains. Functional information was retrieved for all genes using InterProScan [108]. Each InterPro domain was tested for enrichment or depletion in *Ulva* compared with the Volvocales by Fisher's exact tests with a false discovery rate correction (Benjamini-Hochberg FDR method) of 0.05.Interpro domains, proven to be significantly depleted or expanded were grouped at Superfamily-level to reduce redundancy. From TribeMCL [104]gene families, highly conserved families were defined as single copy gene families present in all 20 species (*U. mutabilis*, 3 Chlorophyceae, 6 Trebouxiophyceae, 6 prasinophytes and 4 Streptophyta). For each of the 58 identified single-copy conserved families, protein sequences were aligned using MUSCLE v3.8.31 [109] and alignments concatenated per species. This unedited alignment (42,401 amino acid positions) was used to construct a phylogenetic tree using RAxML v8.2.8 [110] (model PROTGAMMAWAG, 100 bootstraps).

Horizontal gene transfer

To search for HGTs, a blast search *U. mutabilis* proteome (blastp, E-value < 10^{-5}) was carried out against[117] a reduced RefSeq database complemented with data from several algal genomes [118]. The top 1,000 Blastp hits (sorted by bit score) from each query were parsed via custom scripts to extract ≤12 representatives from each phylum to create a taxonomically diverse sample. The blastp hits were re-ordered according to query-hit identity followed by the sampling of another set of representative sequences. The query sequence was then combined with the sets of sampled representative sequences. Sequence alignments were built using Muscle v3.8.31 under default settings [109] followed by trimming using TrimAl v1.2 [111]. FastTree v2.1 [112] was used the under the WAG model to build phylogenetic trees consisting of at least 4 leaves. A custom script was used to sort trees consisting of *U. mutabilis* that was nested among prokaryotes with at least 80% bootstrap supports [117]. Candidate HGT sequences were then reanalyzed using IQtree v1.4.3 [113] with the built-in model selection function, and branch support estimated using ultrafast bootstrap (UFboot) with 1,500 bootstrap replicates (-bb 1,500). Last, we visually confirmed that the putative HGT genes were an integral part of long sequencing reads and flanked up and downstream by non-HGT genes.

Phytohormone bioassay and measurements

Endogenous phytohormones were extracted from thalli (3-weeks-old) and axenic cultures (12.5 mg·mL^{-1}) by the acidified polar solvent 80 % acetonitrile 1 % acetic acid in water (v:v) upon maceration through a tissue lyser (Qiagen, Germany). Compounds were purified to eliminate interferents by multiple steps of solid-phase extractions [119], and quantified by UHPLC-ESI-MS/MS (LC system equipped with a C18-Kinetex column (dimensions: 50 × 1.7 mm, Phenomenex, USA) and coupled to a *Q Exactive* Quadrupole-*Orbitrap* mass spectrometer (Thermo Fisher Scientific, UK). Deuterium-labelled standards (d_5-IAA, d_6-ABA, d_4-SA; ChemIm Czech Republic) were used for quantification. The limit of detection was in the range of 0.8 - 8.0 fmol on the column, whereas the limit of quantification was in the range of 2 - 20 fmol on column depending on the phytohormone. Ethylene emanation was measured by laser photo-acoustics [120].

To test if morphogenesis of *Ulva* into a blade- or tube-like thallus may require specific phytohormones, gametes were inoculated with ABA, GA_3, IAA, JA, SA and ZEA at a final concentration of 10^{-6} and 10^{-9} mol/l. Control treatments consisted of inoculating *Ulva* gametes with bacteria (final optical density 1x10^{-4}). Development was monitored for 2 weeks.

Measurement of DMSP and qPCR.

Expression of DMSP lyases (UM030_0039, UM021_0036), putative S-adenosyl-L-methionine-dependent methyltransferases (UM052_056, UM036_0102) and BCCT transporters (UM033_146, UM033_147, UM033_150, UM033_150) was examined in xenic and axenic 4-5 week-old thalli grown at 8°C or 18°C using qPCR (Supplemental Experimental Procedures: DMSP-qPCR analysis; Table S8). Expression values are based on at least 4 biological replicates. DMSP was measured in xenic thalli grown under both temperatures as outlined in Kessler et al. [79] (n = 12). RNA was extracted from *Ulva mutabilis* (slender strain) using the Spectrum Plant Total RNA Kit (Sigma), including the optional DNaseI digestion, and 500 ng was reverse-transcribed with the PrimeScript RT-PCR Kit (Clontech). qPCR reactions were performed using SYBR Green Master Mix (Clontech) on a BioRad CFX96 Real-Time PCR Detection System. Primers are listed. Amplification efficiencies were determined using LinRegPCR [121] and transcript abundance according to the Pfaffl method [122, 123]. For normalization of expression levels, we used the reference genes UM008_0183 (Ubiquitin) and UM010_0003 (PP2A 65 kDa regulatory subunit A).

DATA AND SOFTWARE AVAILABILITY

Acknowledgments

Research support was provided by NERC grant NBAF925 (to CAM and JHB); UGent Special Research Fund BOF-GOA /01G01715 (to EC), BOF/01J04813 (to ODC, KVP), BOF/01SC2316 (to XL) with infrastructure funded by EMBRC Belgium - FWO project GOH3817N (to ODC); EU Horizon2020 Marie Curie ITN ALFF-Project 642575 (to CMG, ODC, TW, GC, YVDP); EU COST Action FA1406 Phycomorph (to BC, JC, JHB, ODC, SAR and TW); German Research Foundation CRC ChemBioSys 1127 (to TW, JBo, MK, GC, LL); Postdoctoral Fellowship Grant of the Research Foundation – Flanders (to JB-Project 12T3418N); BBSRC-funded MIBTP PhD rotation project to RC and JC; United States Department of Energy DE-EE0003373/001 (to DB); National Science Foundation IGERT for Renewable and Sustainable fuels program 0903675 (to FF); MH is Research Associate of the FNRS.

Author Contributions

J.H.B., O.D.C., T.W., K.V., L.S. and Y.V.D.P. designed research; O.D.C, S.M.K., K.A.B., J.Bl. F.F., M.K., E.V., L.V., E.A., J.Bo., G.C., B.C., R.C., A.D.C., N.F.P., C.M.G., M.H., L.L., F.L., X.L., Z.A.P., M.V.B., P.I.K.W., D.B., J.C.C., S.A.R., D.V.D.S., L.S., T.W. and J.H.B. performed research; S.M.K., K.A.B., M.K., E.V., L.V., E.A., G.C., L.L., X.L., M.V.B., N.F.P., P.I.K.W., L.S., J.H.B. contributed new reagents/analytic tools; O.D.C, S.M.K., K.A.B., J.Bl., F.F., M.K., E.V., L.V., B.C., R.C., A.D.C., M.H., F.L., Z.A.P., J.A.R., D.B., J.C.C., S.A.R., D.V.D.S., A.V., L.S., K.V.P., T.W. and J.H.B. analyzed data; A.D.C., J.B., G.C., S.D., L.L., L.V., and X.L. prepared samples; and O.D.C., S.M.K., K.A.B, F.L., C.A.M., D.B., J.C.C., S.A.R., D.V.D.S., A.V., L.S., K.V., T.W., J.H.B., K.V., Y.V.D.P. wrote the paper.

Declaration Of Interests

The authors declare no competing interests.

References

1. Knoll, A.H. (2011). The multiple origins of complex multicellularity. Annu. Rev. Earth Planet. Sci. *39*, 217-239.
2. Rensing, S.A. (2016). (Why) Does Evolution Favour Embryogenesis? Trends Plant Sci. *21*, 562-573.
3. Sebé-Pedrós, A., Degnan, B.M., and Ruiz-Trillo, I. (2017). The origin of Metazoa: a unicellular perspective. Nat. Rev. Genet. *18*, 498-512.
4. De Smet, I., Voss, U., Lau, S., Wilson, M., Shao, N., Timme, R., Swarup, R., Kerr, I., Hodgman, C., Bock, R., et al. (2010). Unraveling the evolution of auxin signaling. Plant Physiol. *155*, 209-221.
5. Ju, C., Van de Poel, B., Cooper, E.D., Thierer, J.H., Gibbons, T.R., Delwiche, C.F., and Chang, C. (2015). Conservation of ethylene as a plant hormone over 450 million years of evolution. Nature Plants *1*.
6. Herron, M.D., Hackett, J.D., Aylward, F.O., and Michod, R.E. (2009). Triassic origin and early radiation of multicellular volvocine algae. Proc. Natl. Acad. Sci. U. S. A. *106*, 3254-3258.
7. Umen, J.G., and Olson, B.J.S.C. (2012). Genomics of Volvocine Algae. In Genomic insights into the biology of algae, Volume 64, G. Piganeau, ed. (Academic Press), pp. 185-243.
8. Featherston, J., Arakaki, Y., Hanschen, E.R., Ferris, P.J., Michod, R.E., Olson, B.J.S.C., Nozaki, H., and Durand, P.M. (2017). The 4-celled *Tetrabaena socialis* nuclear genome reveals the essential components for genetic control of cell number at the origin of multicellularity in the volvocine lineage. Mol. Biol. Evol., 10.1093/molbev/msx1332.
9. Prochnik, S.E., Umen, J., Nedelcu, A.M., Hallmann, A., Miller, S.M., Nishii, I., Ferris, P., Kuo, A., Mitros, T., Fritz-Laylin, L.K., et al. (2010). Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. Science *329*, 223-226.

10. Hanschen, E.R., Marriage, T.N., Ferris, P.J., Hamaji, T., Toyoda, A., Fujiyama, A., Neme, R., Noguchi, H., Minakuchi, Y., Suzuki, M., et al. (2016). The *Gonium pectorale* genome demonstrates co-option of cell cycle regulation during the evolution of multicellularity. Nat. Commun. *7*, 11370.
11. Leliaert, F., Verbruggen, H., Herron, M., Moreau, H., Smith, D.R., Delwiche, C., and De Clerck, O. (2012). Phylogeny and molecular evolution of the green algae. Crit. Rev. Plant Sci. *31*, 1-46.
12. Mine, I., Menzel, D., and Okuda, K. (2008). Morphogenesis in giant-celled algae. In International Review of Cell and Molecular Biology, Volume 266, K.W. Jeon, ed., pp. 37-83.
13. Cocquyt, E., Verbruggen, H., Leliaert, F., and De Clerck, O. (2010). Evolution and cytological diversification of the green seaweeds (Ulvophyceae). Mol. Biol. Evol. *27*, 2052-2061.
14. Wichard, T., Charrier, B., Mineur, F., Bothwell, J.H., De Clerck, O., and Coates, J.C. (2015). The green seaweed *Ulva*: a model system to study morphogenesis. Front. Plant Sci. *6*, 72.
15. Tan, I.H., Blomster, J., Hansen, G., Leskinen, E., Maggs, C.A., Mann, D.G., Sluimam, H.J., and Stanhope, M.J. (1999). Molecular phylogenetic evidence for a reversible morphogenetic switch controlling the gross morphology of two common genera of green seaweeds, *Ulva* and *Enteromorpha*. Mol. Biol. Evol. *16*, 1011-1018.
16. Hayden, H.S., Blomster, J., Maggs, C.A., Silva, P.C., Stanhope, M.J., and Waaland, J.R. (2003). Linnaeus was right all along: *Ulva* and *Enteromorpha* are not distinct genera. Eur. J. Phycol. *38*, 277-294.
17. Løvlie, A. (1964). Genetic control of division rate and morphogenesis in *Ulva mutabilis* Føyn. C. R. Trav. Lab. Carlsberg *34*, 77-168.
18. Spoerner, M., Wichard, T., Bachhuber, T., Stratmann, J., and Oertel, W. (2012). Growth and thallus morphogenesis of *Ulva mutabilis* (Chlorophyta) depends on a combination of two bacterial species excreting regulatory factors. J. Phycol. *48*, 1433-1447.
19. Ghaderiardakani, F., Coates, J.C., and Wichard, T. (2017). Bacteria-induced morphogenesis of *Ulva intestinalis* and *Ulva mutabilis* (Chlorophyta): a contribution to the lottery theory. FEMS Microbiol. Ecol. *93*.
20. Matsuo, Y., Imagawa, H., Nishizawa, M., and Shizuri, Y. (2005). Isolation of an algal morphogenesis inducer from a marine bacterium. Science *307*, 1598-1598.
21. Oertel, W., Wichard, T., and Weissgerber, A. (2015). Transformation of *Ulva mutabilis* (Chlorophyta) by vector plasmids integrating into the genome. J. Phycol. *51*, 963-979.
22. Hu, C., Li, D., Chen, C., Ge, J., Muller-Karger, F.E., Liu, J., Yu, F., and He, M.X. (2010). On the recurrent *Ulva prolifera* blooms in the Yellow Sea and East China Sea. J. Geophys. Res. *115*, C05017.
23. Smetacek, V., and Zingone, A. (2013). Green and golden seaweed tides on the rise. Nature *504*, 84-88.
24. Korzen, L., Pulidindi, I.N., Israel, A., Abelson, A., and Gedanken, A. (2015). Single step production of bioethanol from the seaweed *Ulva rigida* using sonication. RSC Advances *5*, 16223-16229.
25. Fernand, F., Israel, A., Skjermo, J., Wichard, T., Timmermans, K.R., and Golberg, A. (2016). Offshore macroalgae biomass for bioenergy production: Environmental aspects, technological achievements and challenges. Renewable and Sustainable Energy Reviews.
26. Bikker, P., van Krimpen, M.M., van Wikselaar, P., Houweling-Tan, B., Scaccia, N., van Hal, J.W., Huijgen, W.J.J., Cone, J.W., and López-Contreras, A.M. (2016). Biorefinery of the green seaweed *Ulva lactuca* to produce animal feed, chemicals and biofuels. J. Appl. Phycol. *28*, 3511-3525.
27. Cohen, I., and Neori, A. (1991). *Ulva lactuca* biofilters for marine fishpond effluents. I. Ammonia uptake kinetics and nitrogen content. Bot. Mar. *34*, 475-482.
28. de Paula Silva, P.H., McBride, S., de Nys, R., and Paul, N.A. (2008). Integrating filamentous 'green tide'algae into tropical pond-based aquaculture. Aquaculture *284*, 74-80.

29. Bolton, J.J., Cyrus, M.D., Brand, M.J., Joubert, M., and Macey, B.M. (2016). Why grow *Ulva*? Its potential role in the future of aquaculture. Perspectives in Phycology *3*, 113-120.
30. Zhang, X., Ye, N., Liang, C., Mou, S., Fan, X., Xu, J., Xu, D., and Zhuang, Z. (2012). De novo sequencing and analysis of the *Ulva linza* transcriptome to discover putative mechanisms associated with its successful colonization of coastal ecosystems. BMC Genomics *13*, 565.
31. Yamazaki, T., Ichihara, K., Suzuki, R., Oshima, K., Miyamura, S., Kuwano, K., Toyoda, A., Suzuki, Y., Sugano, S., Hattori, M., et al. (2017). Genomic structure and evolution of the mating type locus in the green seaweed *Ulva partita*. Scientific Reports *7*, 11679.
32. Ranjan, A., Townsley, B.T., Ichihashi, Y., Sinha, N.R., and Chitwood, D.H. (2015). An Intracellular Transcriptomic Atlas of the Giant Coenocyte *Caulerpa taxifolia*. PLoS Genet *11*, e1004900.
33. Føyn, B. (1958). Über die Sexualität und den Generationswechsel von *Ulva mutabilis*. Arch. Protistenkunde *102*, 473-480.
34. Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., and Zdobnov, E.M. (2017). BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol., doi.org/10.1093/molbev/msx1319.
35. Veeckman, E., Ruttink, T., and Vandepoele, K. (2016). Are we there yet? Reliably estimating the completeness of plant genome sequences. The Plant Cell *28*, 1759-1768.
36. Løvlie, A. (1978). On the genetic control of cell cycles during morphogenesis in *Ulva mutabilis*. Dev. Biol. *64*, 164-177.
37. Fjeld, A. (1972). Genetic control of cellular differentiation in Ulva mutabilis. Gene effects in early development. Dev. Biol. *28*, 326-343.
38. Rokas, A. (2008). The molecular origins of multicellular transitions. Curr. Opin. Genet. Dev. *18*, 472-478.
39. Lang, D., Weiche, B., Timmerhaus, G., Richardt, S., Riaño-Pachón, D.M., Corrêa, L.G., Reski, R., Mueller-Roeber, B., and Rensing, S.A. (2010). Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. Genome Biol. Evol. *2*, 488-503.
40. Harashima, H., Dissmeyer, N., and Schnittger, A. (2013). Cell cycle control across the eukaryotic kingdom. Trends Cell Biol. *23*, 345-356.
41. de Mendoza, A., Sebé-Pedrós, A., Šestak, M.S., Matejčić, M., Torruella, G., Domazet-Lošo, T., and Ruiz-Trillo, I. (2013). Transcription factor evolution in eukaryotes and the assembly of the regulatory toolkit in multicellular lineages. Proc. Natl. Acad. Sci. U. S. A. *110*, E4858-E4866.
42. Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierec, J., Samach, A., and Coupland, G. (2006). CONSTANS and the CCAAT Box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. The Plant Cell *18*, 2971-2984.
43. Khanna, R., Kronmiller, B., Maszle, D.R., Coupland, G., Holm, M., Mizuno, T., and Wu, S.-H. (2009). The *Arabidopsis* B-box zinc finger family. The Plant Cell *21*, 3416-3420.
44. Robson, F., Costa, M.M.R., Hepworth, S.R., Vizir, I., Reeves, P.H., Putterill, J., and Coupland, G. (2001). Functional importance of conserved domains in the flowering-time gene CONSTANS demonstrated by analysis of mutant alleles and transgenic plants. Plant J. *28*, 619-631.
45. Crocco, C.D., and Botto, J.F. (2013). BBX proteins in green plants: insights into their evolution, structure, feature and functional diversification. Gene *531*, 44-52.
46. Cockram, J., Thiel, T., Steuernagel, B., Stein, N., Taudien, S., Bailey, P.C., and O'Sullivan, D.M. (2012). Genome dynamics explain the evolution of flowering time CCT domain gene families in the Poaceae. PLoS One *7*, e45307.
47. Putterill, J., Robson, F., Lee, K., Simon, R., and Coupland, G. (1995). The CONSTANS gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell *80*, 847-857.

48. Strayer, C., Oyama, T., Schultz, T.F., Raman, R., Somers, D.E., Más, P., Panda, S., Kreps, J.A., and Kay, S.A. (2000). Cloning of the *Arabidopsis* clock gene TOC1, an autoregulatory response regulator homolog. Science *289*, 768-771.
49. Kaczorowski, K.A., and Quail, P.H. (2003). *Arabidopsis* PSEUDO-RESPONSE REGULATOR7 is a signaling intermediate in phytochrome-regulated seedling deetiolation and phasing of the circadian clock. The Plant Cell *15*, 2654-2665.
50. Liu, J., Shen, J., Xu, Y., Li, X., Xiao, J., and Xiong, L. (2016). Ghd2, a CONSTANS-like gene, confers drought sensitivity through regulation of senescence in rice. J. Exp. Bot. *67*, 5785-5798.
51. Zobell, O., Coupland, G., and Reiss, B. (2005). The family of CONSTANS-like genes in *Physcomitrella patens*. Plant Biology *7*, 266-275.
52. Corellou, F., Schwartz, C., Motta, J.-P., Sanchez, F., and Bouget, F.-Y. (2009). Clocks in the green lineage: comparative functional analysis of the circadian architecture of the picoeukaryote *Ostreococcus*. The Plant Cell *21*, 3436-3449.
53. Serrano, G., Herrera-Palau, R., Romero, J.M., Serrano, A., Coupland, G., and Valverde, F. (2009). *Chlamydomonas* CONSTANS and the evolution of plant photoperiodic signaling. Curr. Biol. *19*, 359-368.
54. Lang, D., and Rensing, S.A. (2015). The Evolution of Transcriptional Regulation in the Viridiplantae and its Correlation with Morphological Complexity. In Evolutionary Transitions to Multicellular Life: Principles and mechanisms, I. Ruiz-Trillo and A.M. Nedelcu, eds. (Dordrecht: Springer Netherlands), pp. 301-333.
55. Zulawski, M., Schulze, G., Braginets, R., Hartmann, S., and Schulze, W.X. (2014). The *Arabidopsis* Kinome: phylogeny and evolutionary insights into functional diversification. BMC Genomics *15*, 548.
56. Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Marechal-Drouard, L., et al. (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science *318*, 245-251.
57. Wheeler, G.L., Miranda-Saavedra, D., and Barton, G.J. (2008). Genome analysis of the unicellular green alga *Chlamydomonas reinhardtii* indicates an ancient evolutionary origin for key pattern recognition and cell-signaling protein families. Genetics *179*, 193-197.
58. Cock, J.M., Sterck, L., Rouze, P., Scornet, D., Allen, A.E., Amoutzias, G., Anthouard, V., Artiguenave, F., Aury, J.-M., Badger, J.H., et al. (2010). The Ectocarpus genome and the independent evolution of multicellularity in the brown algae. Nature, 617.
59. Holzinger, A., Herburger, K., Kaplan, F., and Lewis, L.A. (2015). Desiccation tolerance in the chlorophyte green alga *Ulva compressa*: does cell wall architecture contribute to ecological success? Planta *242*, 477-492.
60. Wolf, S., and Greiner, S. (2012). Growth control by cell wall pectins. Protoplasma *249*, 169-175.
61. Daher, F.B., and Braybrook, S.A. (2015). How to let go: pectin and plant cell adhesion. Front. Plant Sci. *6*, 523.
62. Green, J.L., Kuntz, S.G., and Sternberg, P.W. (2008). Ror receptor tyrosine kinases: orphans no more. Trends Cell Biol. *18*, 536-544.
63. Resnick, D., Pearson, A., and Krieger, M. (1994). The SRCR superfamily: a family reminiscent of the Ig superfamily. Trends Biochem. Sci. *19*, 5-8.
64. Pereira, C.B., Bocková, M., Santos, R.F., Santos, A.M., de Araújo, M.M., Oliveira, L., Homola, J., and Carmo, A.M. (2016). The scavenger receptor SSc5D physically interacts with bacteria through the SRCR-containing N-terminal domain. Front. Immunol. *7*.
65. Bowdish, D.M., and Gordon, S. (2009). Conserved domains of the class A scavenger receptors: evolution and function. Immunol. Rev. *227*, 19-31.

66. Jamet, E., Canut, H., Boudart, G., and Pont-Lezica, R.F. (2006). Cell wall proteins: a new insight through proteomics. Trends Plant Sci. *11*, 33-39.
67. Zimmermann, G., Bäumlein, H., Mock, H.-P., Himmelbach, A., and Schweizer, P. (2006). The multigene family encoding germin-like proteins of barley. Regulation and function in basal host resistance. Plant Physiol. *142*, 181-192.
68. Provasoli, L. (1958). Effect of plant hormones on *Ulva*. Biol. Bull. *114*, 375-384.
69. Gupta, V., Kumar, M., Brahmbhatt, H., Reddy, C., Seth, A., and Jha, B. (2011). Simultaneous determination of different endogenetic plant growth regulators in common green seaweeds using dispersive liquid–liquid microextraction method. Plant Physiol. Biochem. *49*, 1259-1263.
70. Wang, C., Liu, Y., Li, S.-S., and Han, G.-Z. (2015). Insights into the origin and evolution of the plant hormone signaling machinery. Plant Physiol. *167*, 872-886.
71. Amin, S.A., Hmelo, L.R., van Tol, H.M., Durham, B.P., Carlson, L.T., Heal, K.R., Morales, R.L., Berthiaume, C.T., Parker, M.S., Djunaedi, B., et al. (2015). Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. Nature *522*, 98-101.
72. Le Bail, A., Billoud, B., Kowalczyk, N., Kowalczyk, M., Gicquel, M., Le Panse, S., Stewart, S., Scornet, D., Cock, J.M., Ljung, K., et al. (2010). Auxin metabolism and function in the multicellular brown alga *Ectocarpus siliculosus*. Plant Physiol. *153*, 128-144.
73. Joint, I., Tait, K., Callow, M.E., Callow, J.A., Milton, D., Williams, P., and Cámara, M. (2002). Cell-to-cell communication across the prokaryote-eukaryote boundary. Science *298*, 1207-1207.
74. Tait, K., Joint, I., Daykin, M., Milton, D.L., Williams, P., and Camara, M. (2005). Disruption of quorum sensing in seawater abolishes attraction of zoospores of the green alga *Ulva* to bacterial biofilms. Environ. Microbiol. *7*, 229-240.
75. Sung, M.-S., Hsu, Y.-T., Hsu, Y.-T., Wu, T.-M., and Lee, T.-M. (2009). Hypersalinity and hydrogen peroxide upregulation of gene expression of antioxidant enzymes in Ulva fasciata against oxidative stress. Mar. Biotechnol. *11*, 199.
76. Passardi, F., Penel, C., and Dunand, C. (2004). Performing the paradoxical: how plant peroxidases modify the cell wall. Trends Plant Sci. *9*, 534-540.
77. Foflonker, F., Price, D.C., Qiu, H., Palenik, B., Wang, S., and Bhattacharya, D. (2015). Genome of the halotolerant green alga *Picochlorum* sp. reveals strategies for thriving under fluctuating environmental conditions. Environ. Microbiol. *17*, 412-426.
78. Wichard, T. (2016). Identification of metallophores and organic ligands in the chemosphere of the marine macroalga *Ulva* (Chlorophyta) and at land-sea interfaces. Front. Mar. Sci. *3*.
79. Kessler, R.W., Weiss, A., Kuegler, S., Hermes, C., and Wichard, T. (2018). Macroalgal-bacterial interactions: Role of dimethylsulfoniopropionate in microbial gardening by *Ulva* (Chlorophyta). Mol. Ecol., 10.1111/mec.14472.
80. Gage, D.A., Rhodes, D., Nolte, K.D., and Hicks, W.A. (1997). A new route for synthesis of dimethylsulphoniopropionate in marine algae. Nature *387*, 891.
81. Summers, P.S., Nolte, K.D., Cooper, A.J., Borgeas, H., Leustek, T., Rhodes, D., and Hanson, A.D. (1998). Identification and stereospecificity of the first three enzymes of 3-dimethylsulfoniopropionate biosynthesis in a chlorophyte alga. Plant Physiol. *116*, 369-378.
82. Curson, A.R., Liu, J., Martínez, A.B., Green, R.T., Chan, Y., Carrión, O., Williams, B.T., Zhang, S.-H., Yang, G.-P., and Page, P.C.B. (2017). Dimethylsulfoniopropionate biosynthesis in marine bacteria and identification of the key gene in this process. Nat. Microbiol. *2*, 17009.
83. Curson, A.R.J., Williams, B.T., Pinchbeck, B.J., Sims, L.P., Martínez, A.B., Rivera, P.P.L., Kumaresan, D., Mercadé, E., Spurgin, L.G., Carrión, O., et al. (2018). DSYB catalyses the key step of dimethylsulfoniopropionate biosynthesis in many phytoplankton. Nat. Microbiol.

84. Lyon, B.R., Lee, P.A., Bennett, J.M., DiTullio, G.R., and Janech, M.G. (2011). Proteomic analysis of a sea-ice diatom: salinity acclimation provides new insight into the dimethylsulfoniopropionate production pathway. Plant Physiol. *157*, 1926-1941.
85. Sun, L., Curson, A.R., Todd, J.D., and Johnston, A.W. (2012). Diversity of DMSP transport in marine bacteria, revealed by genetic analyses. Biogeochemistry, 121-130.
86. Todd, J.D., Curson, A.R., Nikolaidou-Katsaraidou, N., Brearley, C.A., Watmough, N.J., Chan, Y., Page, P.C., Sun, L., and Johnston, A.W. (2010). Molecular dissection of bacterial acrylate catabolism–unexpected links with dimethylsulfoniopropionate catabolism and dimethyl sulfide production. Environ. Microbiol. *12*, 327-343.
87. Vila-Costa, M., Simó, R., Harada, H., Gasol, J.M., Slezak, D., and Kiene, R.P. (2006). Dimethylsulfoniopropionate uptake by marine phytoplankton. Science *314*, 652-654.
88. Spielmeyer, A., Gebser, B., and Pohnert, G. (2011). Investigations of the uptake of dimethylsulfoniopropionate by phytoplankton. ChemBioChem *12*, 2276-2279.
89. Ito, T., Asano, Y., Tanaka, Y., and Takabe, T. (2011). Regulation of biosynthesis of dimethylsulfoniopropionate and its uptake in sterile mutant of *Ulva pertusa* (Chlorophyta). J. Phycol. *47*, 517-523.
90. Alcolombri, U., Ben-Dor, S., Feldmesser, E., Levin, Y., Tawfik, D.S., and Vardi, A. (2015). Identification of the algal dimethyl sulfide–releasing enzyme: A missing link in the marine sulfur cycle. Science *348*, 1466-1469.
91. Lovelock, J.E., Maggs, R., and Rasmussen, R. (1972). Atmospheric dimethyl sulphide and the natural sulphur cycle. Nature *237*, 452-453.
92. Charlson, R.J., Lovelock, J.E., Andreaei, M.O., and Warren, S.G. (1987). Oceanic phytoplankton, atmospheric sulphur, cloud. Nature *326*, 655-661.
93. de Souza, M.P., Chen, Y.P., and Yoch, D.C. (1996). Dimethylsulfoniopropionate lyase from the marine macroalga *Ulva curvata*: purification and characterization of the enzyme. Planta *199*, 433-438.
94. Steinke, M., and Kirst, G.O. (1996). Enzymatic cleavage of dimethylsulfoniopropionate (DMSP) in cell-free extracts of the marine macroalga *Enteromorpha clathrata* (Roth) Grev.(Ulvales, Chlorophyta). J. Exp. Mar. Biol. Ecol. *201*, 73-85.
95. Coneva, V., and Chitwood, D.H. (2015). Plant architecture without multicellularity: quandaries over patterning and the soma-germline divide in siphonous algae. Front. Plant Sci. *6*.
96. Wichard, T., and Oertel, W. (2010). Gametogenesis and gamete release of *Ulva mutabilis* and *Ulva lactuca* (Chlorophyta): Regulatory effects and chemical characterization of the "swarming inhibitor". J. Phycol. *46*, 248-259.
97. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. *27*, 722-736.
98. Zimin, A.V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L., and Yorke, J.A. (2013). The MaSuRCA genome assembler. Bioinformatics *29*, 2669-2677.
99. Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M.-F., Lió, P., Crescenzi, P., Fani, R., and Fondi, M. (2015). MeDuSa: a multi-draft based scaffolder. Bioinformatics *31*, 2443-2451.
100. Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. (2010). Scaffolding pre-assembled contigs using SSPACE. Bioinformatics *27*, 578-579.
101. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., and Young, S.K. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PloS one *9*, e112963.
102. Smit, A., Hubley, R., and Green, P. (2013–2015). RepeatMasker Open-4.0 [Internet]. Institute for Systems Biology. Available from: www.repeatmasker.org.

103. Vandepoele, K., Van Bel, M., Richard, G., Van Landeghem, S., Verhelst, B., Moreau, H., Van de Peer, Y., Grimsley, N., and Piganeau, G. (2013). pico-PLAZA, a genome database of microbial photosynthetic eukaryotes. Environ. Microbiol. *15*, 2147-2153.
104. Enright, A., Van Dongen, S., and Ouzounis, C. (2004). TribeMCL: An efficient algorithm for large scale detection of protein families. Retrieved November *21*, 2005.
105. Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. *13*, 2178-2189.
106. Proost, S., Fostier, J., De Witte, D., Dhoedt, B., Demeester, P., Van de Peer, Y., and Vandepoele, K. (2011). i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. Nucleic Acids Res. *40*, e11-e11.
107. Felsenstein, J. (2005). PHYLIP: Phylogenetic inference program, version 3.6. University of Washington, Seattle.
108. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., and Nuka, G. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236-1240.
109. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792-1797.
110. Stamatakis, A. (2014). RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics *30*, 1312-1313.
111. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972-1973.
112. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. PloS one *5*, e9490.
113. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2014). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268-274.
114. Føyn, B. (1959). Geschlechtskontrollierte Vererbung bei der marinen Grünalge *Ulva mutabilis*. Arch. Protistenkunde *104*, 236-253.
115. Stratmann, J., Paputsoglu, G., and Oertel, W. (1996). Differentiation of *Ulva mutabilis* (Chlorophyta) gametangia and gamete release are controlled by extracellular inhibitors. J. Phycol. *32*, 1009-1021.
116. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.
117. Qiu, H., Cai, G., Luo, J., Bhattacharya, D., and Zhang, N. (2016). Extensive horizontal gene transfers between plant pathogenic fungi. BMC biology *14*, 41.
118. Qiu, H., Price, D.C., Weber, A.P., Reeb, V., Yang, E.C., Lee, J.M., Kim, S.Y., Yoon, H.S., and Bhattacharya, D. (2013). Adaptation through horizontal gene transfer in the cryptoendolithic red alga Galdieria phlegrea. Curr. Biol. *23*, R865-R866.
119. Matsuura, T., Mori, I.C., Ikeda, Y., Hirayama, T., and Mikami, K. (2018). Comprehensive phytohormone quantification in the red alga Pyropia yezoensis by liquid chromatography–mass spectrometry. In Protocols for Macroalgae Research, B. Charrier, T. Wichard and C.R.K. Reddy, eds. (Boca Raton: CRC Press, Francis & Taylor Group).
120. Hu, Y., Depaepe, T., Smet, D., Hoyerova, K., Klíma, P., Cuypers, A., Cutler, S., Buyst, D., Morreel, K., and Boerjan, W. (2017). ACCERBATIN, a small molecule at the intersection of auxin and reactive oxygen species homeostasis with herbicidal properties. J. Exp. Bot. *68*, 4185-4203.
121. Ruijter, J., Ramakers, C., Hoogaars, W., Karlen, Y., Bakker, O., Van den Hoff, M., and Moorman, A. (2009). Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res. *37*, e45-e45.

122. Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Res. *29*, e45-e45.
123. Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome biology *3*, research0034. 0031.

Figure Legends

Figure 1 | A. *Ulva mutabilis* external morphology of a blade (left) and tubular thallus (right). Both growth forms are part of the naturally occurring morphological variation (bar = 1 cm). B. Under axenic conditions *Ulva* develops into a callus-like morphology without cell differentiation. Typical protrusions of the malformed exterior cell wall are visible (bar = 100 µm). C, D. *Ulva* can be easily cultured under standardized conditions associated with two bacterial strains, *Roseovarius* sp. MS2 and *Maribacter* sp. MS6, in batch cultures (C) or tissue flasks (D) (bar = 1 cm).

Figure 2 | Comparison of genome size and number of protein coding genes among green algal genomes. Color coding indicates classes.

Figure 3 | Predicted pattern of gain and loss of gene families during the evolution of green algae and land plants. The number of gene families acquired or lost (values indicated in blue along each branch in the tree) was estimated using the Dollo parsimony principle (see Methods). For each species, the total number of gene families, the number of orphans (genes that lack homologues in the eukaryotic data set) and the number of genes are indicated, as well as habitat and morphological characteristics. Maximum likelihood bootstrap values are indicated in black at each node.

Figure 4 | Comparative analysis of transcription-associated proteins. A. Heatmap of transcription factors comparing *Ulva* with a selection of green algae (*Bathycoccus prasinos*, bpr; *Chlamydomonas reinhardtii*, cre; *Chlorella variabilis*, CN64a; *Gonium pectorale*, gpe; *Micromonas pusilla*, mpu; *Micromonas* sp., m299; *Ostreococcus lucimarinus*, olu; *O. tauri*, ota and o809; *Picochlorum* sp. SE3, pse3; *Volvox carteri*, vca), streprophytes (*Klebsormidium nitens*, kni), land plants (*Arabidopsis thaliana*, ath; *Oryza sativa*, osa; *Physcomitrella patens*, ppa) and red algae (*Chondrus cripus*, ccr; *Cyanidioschyzon merolae*, cme) (see Suppl. Data1). B. Maximum likelihood phylogeny of CO-like transcription factors which are expanded in *Ulva*. Roman numbers refer to the classification as in Khanna et al. [38]. C. Examples of tandem distributions of *Ulva* CO-like genes (containing a CCT and B-box domain), and genes containing either a CCT or B-box domain on contig 003, 053 and 154.

Figure 5 | Comparative analysis of enriched and depleted InterPro domains in *Ulva mutabilis* relative to *Chlamydomonas reinhardtii, Volvox carteri* and/or *Gonium pectorale* (Fisher's exact test, FDR corrected P-value < 0.05). Significant differences are denoted with squares if significant in *Ulva* and *Caulerpa*, circles if significant in *Ulva* only. Z-scores represent number of IPR hits normalized by the total number of hits per species. Abbreviations see Suppl. Table 7.

Figure 6 | Overview of phytohormone biosynthesis pathways and distribution of the biosynthetic enzymes across the Streptophyta and Chlorophyta lineages. (A) Current models of the biosynthesis

pathways for the phytohormones ABA (abscisic acid), BR (brassinosteroids), CK (cytokinins), ET (ethylene), GA (gibberillins), IAA (auxin), JA (jasmonic acid), SA (salicylic acid), and SL (strigolactones). (B) Presence of putative homologs/orthologs of the main biosynthetic enzymes in 20 species based on the pico-PLAZA gene families and subfamilies. Phytohormones shown in green rectangles were identified in both axenic and non-axenic cultures of *Ulva mutabilis* (Table S5). Abbreviations: ZEP, zeaxanthin eioxidase; NCED, nine-cis-epoxycarotenoid deoxygenate; SDR, short-chain dehydrogenase reductase; AAO, aldehyde oxidase; DET2, deetiolated2 (steroid 5alfa reductase); DWF4, dwarf4 (CYP90B); CPD, constitutive photomorphogenic dwarf (CYP90A); IPT, isopentenyltransferase; LOG, lonely guy (lysine decarboxylase); SAMS, s-adenosyl methionine synthetase; ACS, ACC synthase; ACO, ACC oxidase; CPS, CDP/*ent*-kaurene synthase; KO, *ent*-kaurene oxidase; KAO, *ent*-kaurenoic acid oxidase; TAA, tryptophan aminotransferase; YUC, YUCCA (flavin monooxygenase); AMI, amidase; NIT, nitrilase; LOX, lipoxygenase; AOS, allene oxide synthase; AOC, allene oxide cyclase; OPR, oxo-phytodienoate reductase; ICS, isochorismate synthase; PAL, phenylalanine ammonia-lyase; D27, dwarf27 (all-trans/9-cis-B-carotene isomerase); CCD, carotenoid cleavage dioxygenase; MAX1, more axillary branches1 (CYP711A). For complete list of abbreviations see Suppl.Data1.

Figure 7 | Biosynthesis, transport and catabolism of DMSP. A. Maintenance of DMSP concentration in *Ulva* is a combination of *de novo* synthesis by conversion of methionine (after [70]), import from the environment [79] putatively using BCCT transporters, and degradation into acrylate and DMS by DMSP lyases. B. Expression analysis (qPCR) of the Alma DMSP lyase homologues, putative S-adenosyl-L-methionine-dependent methyltransferases and BCCT transporters in xenic and axenic *Ulva* at 18□C and 8□C. Results are shown are based on 4 replicates, except UM030_0039 for which 8 replicates were used. Significant expression values (Fisher's post hoc test) are indicated for each gene. C. Maximum likelihood phylogeny of the DMSP lyase, indicating lateral gene transfer of the Alma gene from the 'chromalveolate' lineage.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
TruSeq DNA PCR-Free Library Preparation Kit	Illumina	Catalog #: FC-121-3001
Nextera Mate Pair Sample Preparation Kit	Illumina	Catalog #: FC-132-1001
ScriptSeq v2 RNA-Seq Library Preparation Kit	Illumina	Catalog #: SSV21106
Deposited Data		
Raw and analyzed data	This paper	ENA: PRJEB25750
Experimental Models: Organisms/Strains		
Ulva mutabilis Føyn (wildtype and slender strains)	[96]	Friedrich Schiller University Jena, Germany
Roseovarius sp. MS2	[18]	Friedrich Schiller University Jena, Germany
Maribacter sp. MS6	[18]	Friedrich Schiller University Jena, Germany
Software and Algorithms		
Canu (v1.6)	[97]	https://github.com/marbl/canu
MaSuRCA (v3.2.3)	[98]	https://github.com/alekseyzimin/masurca
MEDUSA	[99]	https://github.com/combogenomics/medusa
SSPACE (v3.0)	[100]	https://github.com/nsoranzo/sspace_basic
Pilon (v1.20)	[101]	https://github.com/broadinstitute/pilon
RepeatMasker (v4.0.7)	[102]	https://github.com/rmhubley/RepeatMasker
BUSCO (v3.0.2b)	[34]	https://gitlab.com/ezlab/busco

Core gene family analysis	[35]	ftp://ftp.psb.ugent.be/pub/plaza/plaza_public_02_5/coreGF
Pico-PLAZA (v2.0)	[103]	https://bioinformatics.psb.ugent.be/plaza/versions/pico-plaza/
TribeMCL	[104]	https://micans.org/mcl/
OrthoMCL	[105]	http://orthomcl.org/orthomcl/
i-ADHoRe (v3.0.01)	[106]	http://bioinformatics.psb.ugent.be/beg/tools/i-adhore30
PHYLIP (Dollop)	[107]	http://evolution.genetics.washington.edu/phylip.html
InterProScan	[108]	https://www.ebi.ac.uk/interpro/
Muscle (v3.8.31)	[109]	https://www.ebi.ac.uk/Tools/msa/muscle/
RAxML (v8.2.4)	[110]	https://sco.h-its.org/exelixis/web/software/raxml/index.html
TrimAl (v1.2)	[111]	http://trimal.cgenomics.org/
FastTree (v2.1.7)	[112]	http://www.microbesonline.org/fasttree/
IQtree (v1.4.3)	[113]	http://www.iqtree.org/
Other		
Assembly and annotation data	This paper	ORCAE - http://bioinformatics.psb.ugent.be/orcae/overview/Ulvmu

Figure 1

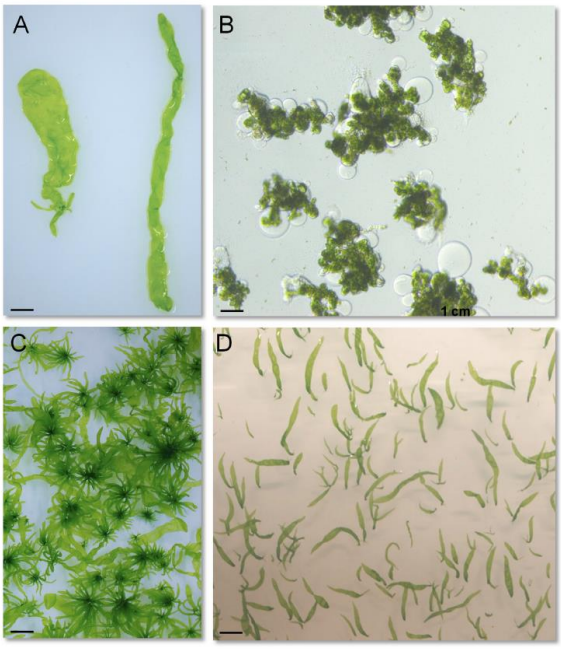


Figure 2

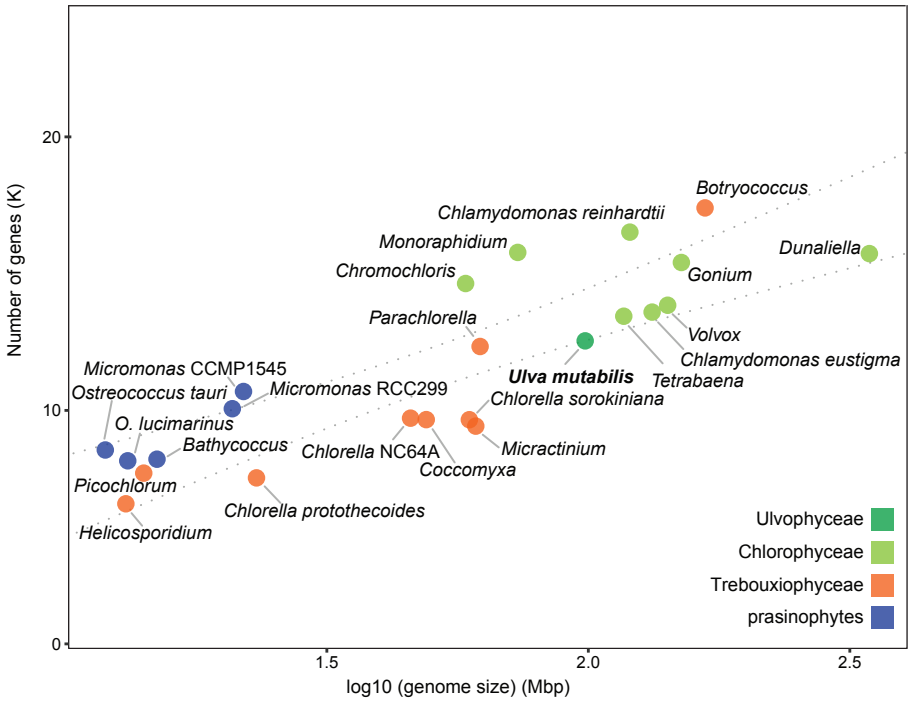

Number of genes (K)
20
10
0
1.5
2.0
2.5
log10 (genome size) (Mbp)
Botryococcus
Chlamydomonas reinhardtii
Monoraphidium
Dunaliella
Chromochloris
Gonium
Parachlorella
Volvox
Chlamydomonas eustigma
Micromonas CCMP1545
Ostreococcus tauri
Micromonas RCC299
Ulva mutabilis
Tetrabaena
Chlorella sorokiniana
O. lucimarinus
Bathycoccus
Chlorella NC64A
Micractinium
Coccomyxa
Picochlorum
Chlorella protothecoides
Helicosporidium
Ulvophyceae
Chlorophyceae
Trebouxiophyceae
prasinophytes

Figure 3

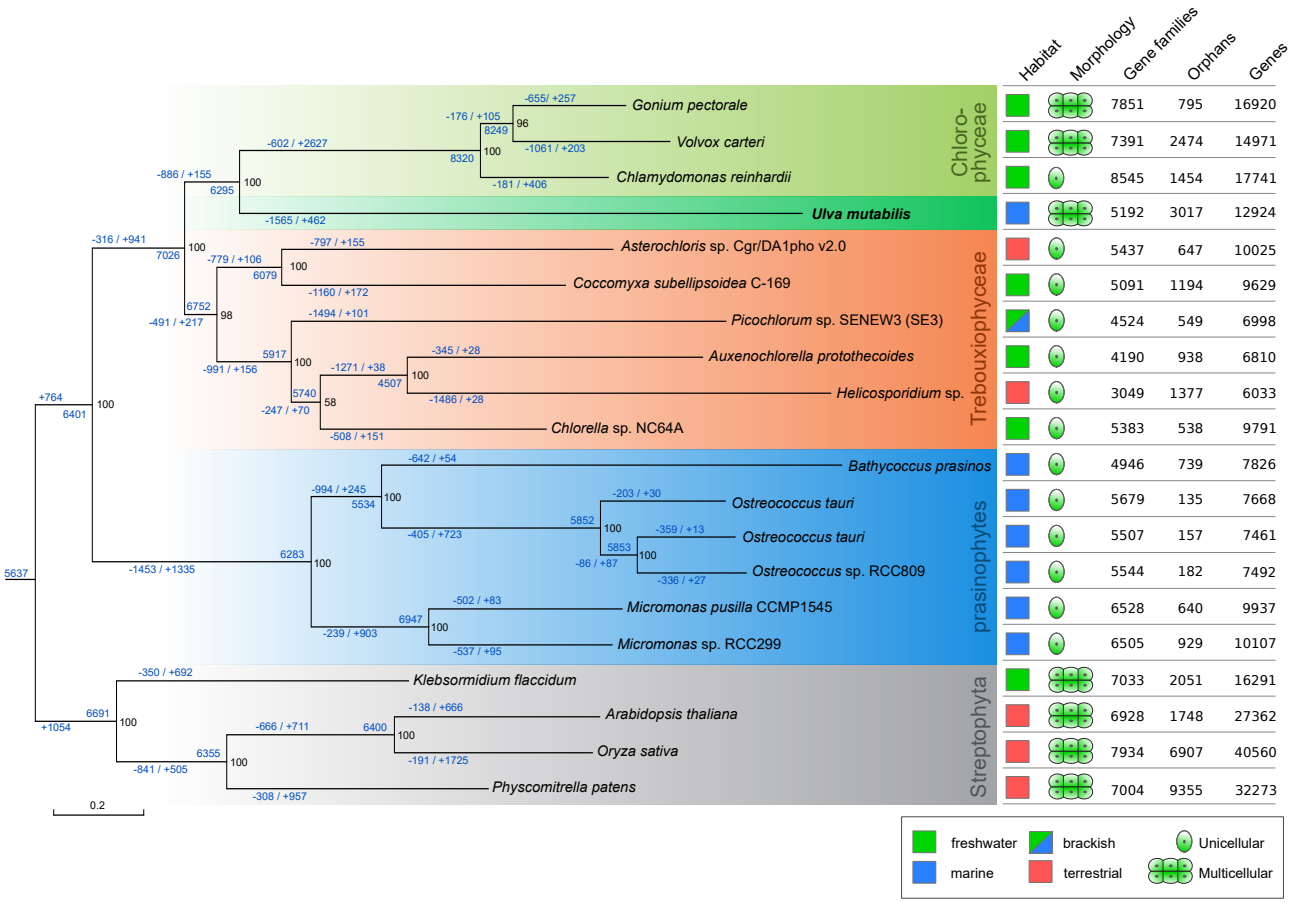
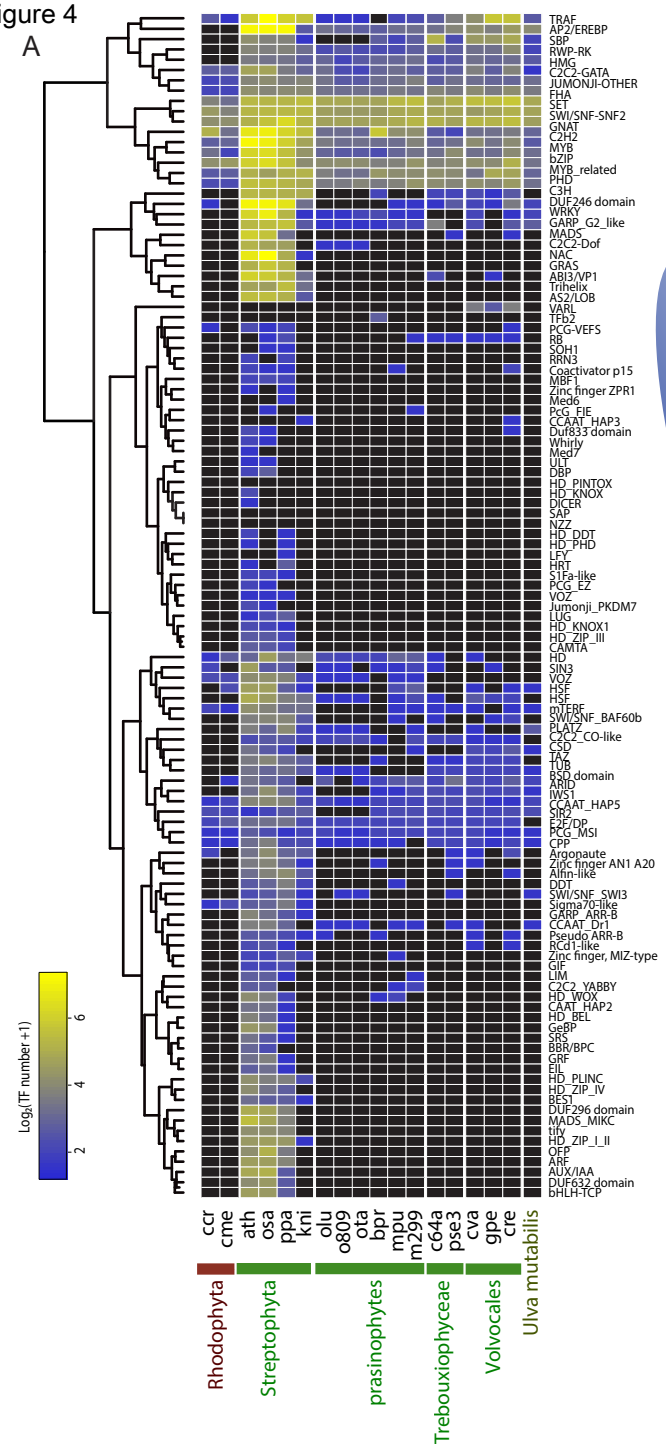
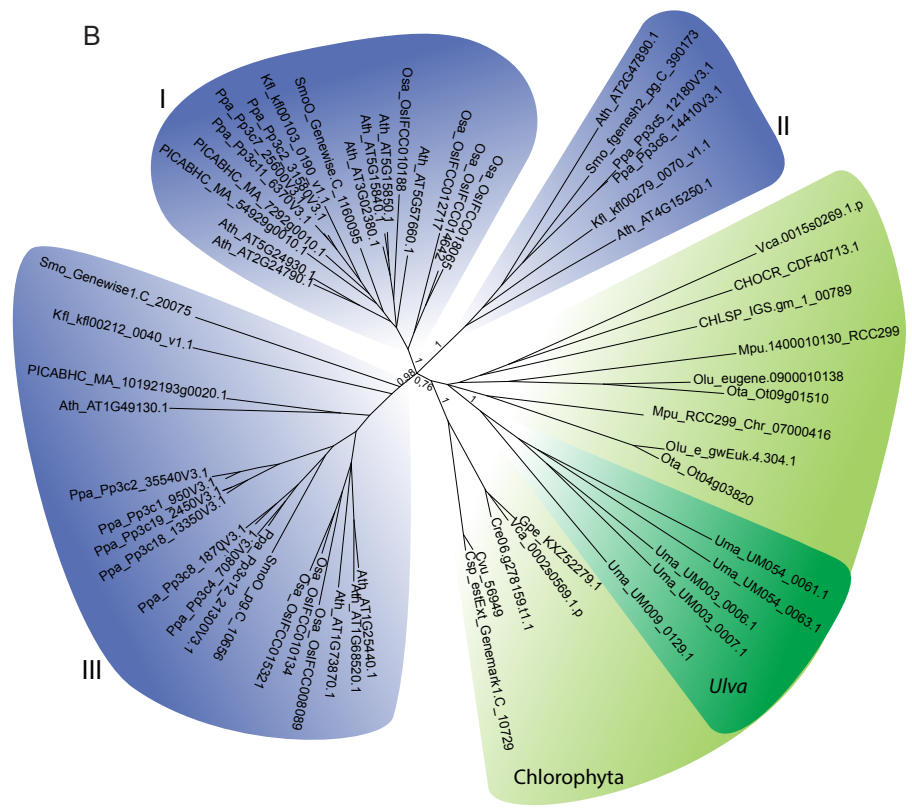

Habitat
Morphology
Gene families
Orphans
Genes
Chloro-phyceae
Trebouxiophyceae
prasinophytes
Streptophyta
Gonium pectorale
Volvox carteri
Chlamydomonas reinhardii
Ulva mutabilis
Asterochloris sp. Cgr/DA1pho v2.0
Coccomyxa subellipsoidea C-169
Picochlorum sp. SENEW3 (SE3)
Auxenochlorella protothecoides
Helicosporidium sp.
Chlorella sp. NC64A
Bathycoccus prasinos
Ostreococcus tauri
Ostreococcus tauri
Ostreococcus sp. RCC809
Micromonas pusilla CCMP1545
Micromonas sp. RCC299
Klebsormidium flaccidum
Arabidopsis thaliana
Oryza sativa
Physcomitrella patens
0.2
7851 795 16920
7391 2474 14971
8545 1454 17741
5192 3017 12924
5437 647 10025
5091 1194 9629
4524 549 6998
4190 938 6810
3049 1377 6033
5383 538 9791
4946 739 7826
5679 135 7668
5507 157 7461
5544 182 7492
6528 640 9937
6505 929 10107
7033 2051 16291
6928 1748 27362
7934 6907 40560
7004 9355 32273
freshwater
marine
brackish
terrestrial
Unicellular
Multicellular

Figure 4


A
TRAF
AP2/EREBP
SBP
RWP-RK
HMG
C2C2-GATA
JUMONJI-OTHER
FHA
SET
SWI/SNF-SNF2
GNAT
C2H2
MYB
bZIP
MYB_related
PHD
C3H
DUF246 domain
WRKY
GARP_G2_like
MADS
C2C2-Dof
NAC
GRAS
ABI3/VP1
Trihelix
AS2/LOB
VARL
TFb2
PCG-VEFS
SOH1
RRN3
Coactivator p15
MBF1
Zinc finger ZPR1
Med6
PcG_EIE
CCAAT_HAP3
Duf833 domain
Whirly
Med7
ULT
DBP
HD_PINTOX
HD_KNOX
DICER
SAP
NZZ
HD_DDT
HD_PHD
LFY
HRT
S1Fa-like
PCG_EZ
VOZ
Jumonji_PKDM7
LUG
HD_KNOX1
HD_ZIP_III
CAMTA
HD
SIN3
VOZ
HSF
HSF
mTERF
SWI/SNF_BAF60b
PLATZ
C2C2_CO-like
CSD
TAZ
TUB
B3D domain
ARID
IWS1
CCAAT_HAP5
SRF
E2F/DP
PCG_MSI
CPP
Argopaute
Zinc finger AN1 A20
Alfin-like
DDT
SWI/SNF_SWI3
Sigma70-like
GARP_ARR-B
CCAAT_Dr1
Pseudo ARR-B
RCd1-like
Zinc finger, MIZ-type
GIF
LIM
C2C2_YABBY
HD_WOX
CAAT_HAP2
HD_BEL
GeBP
SRS
BBR/BPC
GRF
EIL
HD_PLINC
HD_ZIP_IV
BES1
DUF296 domain
MADS_MIKC
tify
HD_ZIP_I_II
OFP
ARF
AUX/IAA
DUF632 domain
bHLH-TCP
Log2(TF number + 1)
2
4
6
ccr
cme
ath
osa
ppa
kni
olu
o809
ota
bpr
mpu
m299
c64a
pse3
cva
gpe
cre
Ulva mutabilis
Rhodophyta
Streptophyta
prasinophytes
Trebouxiophyceae
Volvocales


B
I
II
III
Ulva
Chlorophyta
Osa_OsIFCO101882
Osa_OsIFCO101883
Osa_OsIFCO127177
Osa_OsIFCO101885
Ath_AT5G57660.1
Ath_AT5G15850.1
Ath_AT5G15840.1
Ath_AT3G02360.1
Smo_Genewise1.C_1160095
SmoO_Genewise.C_1190.v1.1
Kfl_kfl00103_0190.v1.1
Ppa_Pp3c2_31580V3.1
Ppa_Pp3c7_25800V3.1
PICABHC_MA_729300g0010.1
PICABHC_MA_54929g0010.1
Ath_AT5G24930.1
Ath_AT2G24790.1
Ath_AT2G47890.1
Smo_fgenesh2_pg.C_390173
Ppa_Pp3c5_12180V3.1
Ppa_Pp3c6_14410V3.1
Kfl_kfl00279_0070_v1.1
Ath_AT4G15250.1
Vca.0015s0269.1.p
CHOCR_CDF40713.1
CHLSP_IGS.gm_1_00789
Mpu.1400010130_RCC299
Olu_eugene.0900010138
Ota_Ot09g01510
Mpu_RCC299_Chr_07000416
Olu_e_gwEuk.4.304.1
Ota_Ot04g03820
Uma_UM054_0061.1
Uma_UM054_0063.1
Uma_UM003_0006.1
Uma_UM003_0007.1
Uma_UM009_0129.1
Gpe_KXZ52279.1
Vca_0002s0569.1.p
Cre06.g278159.t1.1
Cvu_56949
Csl_zsExt_Genemark1.C_10729
Smo_Genewise1.C_20075
Kfl_kfl00212_0040_v1.1
PICABHC_MA_10192193g0020.1
Ath_AT1G49130.1
Ppa_Pp3c2_35540V3.1
Ppa_Pp3c1_950V3.1
Ppa_Pp3c19_2450V3.1
Ppa_Pp3c18_13350V3.1
Ppa_Pp3c8_1870V3.1
Ppa_Pp3c4_7080V3.1
Ppa_Pp3c12_21900V3.1
Smoo_pg.C_10656
Osa_OsIFCO101534
Osa_OsIFCO101536
Osa_OsIFCO101537
Osa_OsIFCO008089
Ath_AT1G25440.1
Ath_AT1G68520.1
Ath_AT1G73870.1
1
0.98
0.76

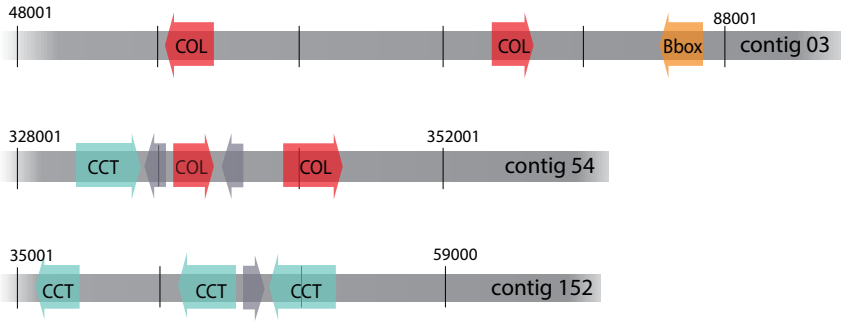

C
48001
88001
COL
COL
Bbox
contig 03
328001
352001
CCT
COL
COL
contig 54
35001
59000
CCT
CCT
CCT
contig 152

Figure 5

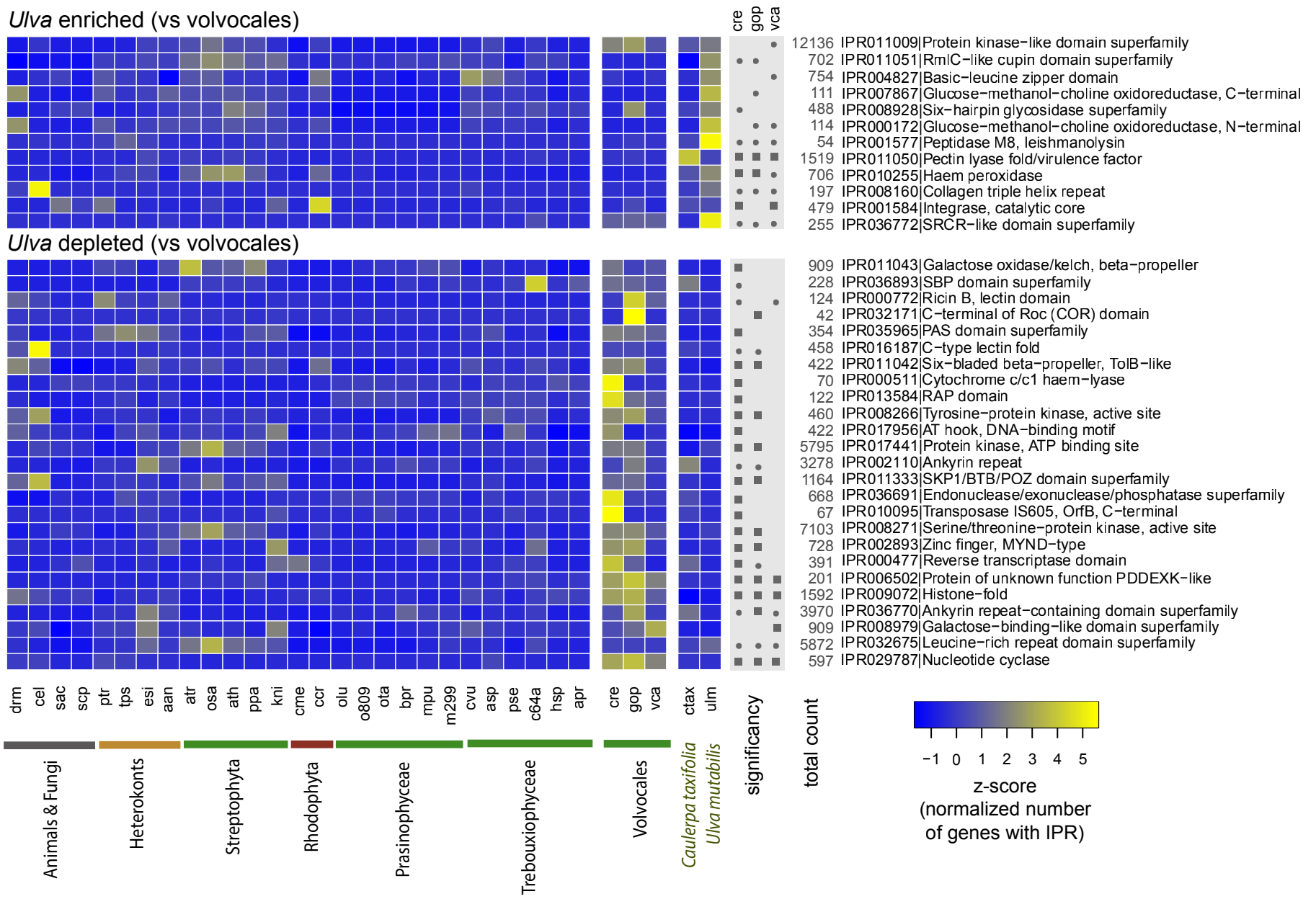

Ulva enriched (vs volvocales)
Ulva depleted (vs volvocales)
cre gop vca
12136 IPR011009|Protein kinase−like domain superfamily
702 IPR011051|RmlC−like cupin domain superfamily
754 IPR004827|Basic−leucine zipper domain
111 IPR007867|Glucose−methanol−choline oxidoreductase, C−terminal
488 IPR008928|Six−hairpin glycosidase superfamily
114 IPR000172|Glucose−methanol−choline oxidoreductase, N−terminal
54 IPR001577|Peptidase M8, leishmanolysin
1519 IPR011050|Pectin lyase fold/virulence factor
706 IPR010255|Haem peroxidase
197 IPR008160|Collagen triple helix repeat
479 IPR001584|Integrase, catalytic core
255 IPR036772|SRCR−like domain superfamily
909 IPR011043|Galactose oxidase/kelch, beta−propeller
228 IPR036893|SBP domain superfamily
124 IPR000772|Ricin B, lectin domain
42 IPR032171|C−terminal of Roc (COR) domain
354 IPR035965|PAS domain superfamily
458 IPR016187|C−type lectin fold
422 IPR011042|Six−bladed beta−propeller, TolB−like
70 IPR000511|Cytochrome c/c1 haem−lyase
122 IPR013584|RAP domain
460 IPR008266|Tyrosine−protein kinase, active site
422 IPR017956|AT hook, DNA−binding motif
5795 IPR017441|Protein kinase, ATP binding site
3278 IPR002110|Ankyrin repeat
1164 IPR011333|SKP1/BTB/POZ domain superfamily
668 IPR036691|Endonuclease/exonuclease/phosphatase superfamily
67 IPR010095|Transposase IS605, OrfB, C−terminal
7103 IPR008271|Serine/threonine−protein kinase, active site
728 IPR002893|Zinc finger, MYND−type
391 IPR000477|Reverse transcriptase domain
201 IPR006502|Protein of unknown function PDDEXK−like
1592 IPR009072|Histone−fold
3970 IPR036770|Ankyrin repeat−containing domain superfamily
909 IPR008979|Galactose−binding−like domain superfamily
5872 IPR032675|Leucine−rich repeat domain superfamily
597 IPR029787|Nucleotide cyclase
drm cel sac scp ptr tps esi aan atr osa ath ppa kni cme ccr olu o809 ota bpr mpu m299 cvu asp pse c64a hsp apr cre gop vca ctax ulm
Animals & Fungi
Heterokonts
Streptophyta
Rhodophyta
Prasinophyceae
Trebouxiophyceae
Volvocales
Caulerpa taxifolia
Ulva mutabilis
significancy
total count
−1 0 1 2 3 4 5
z-score
(normalized number
of genes with IPR)

Figure 6

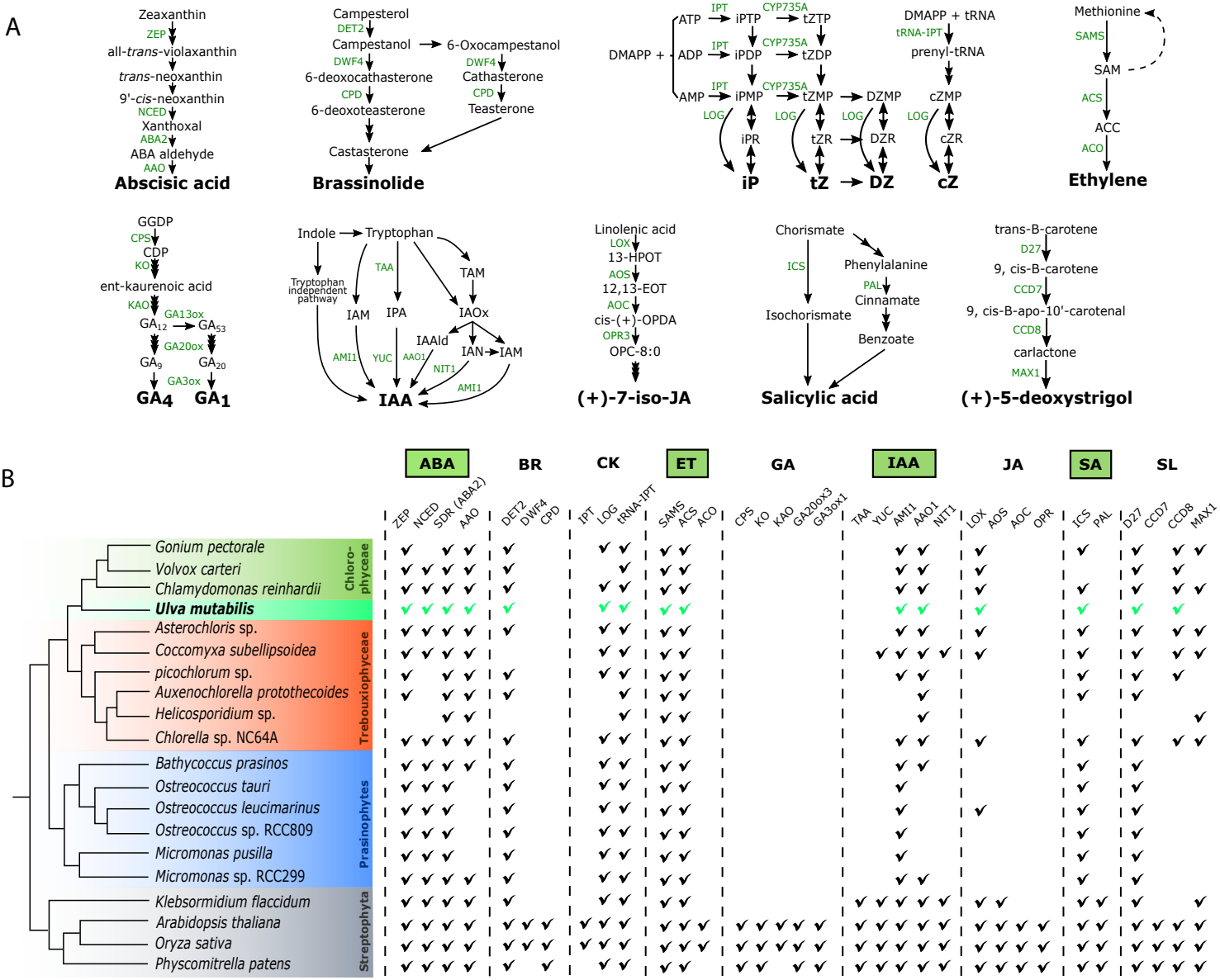

A
Zeaxanthin
ZEP
all-trans-violaxanthin
trans-neoxanthin
9'-cis-neoxanthin
NCED
Xanthoxal
ABA2
ABA aldehyde
AAO
Abscisic acid
Campesterol
DET2
Campestanol
6-Oxocampestanol
DWF4
6-deoxocathasterone
Cathasterone
CPD
6-deoxoteasterone
Teasterone
Castasterone
Brassinolide
DMAPP +
ATP
ADP
AMP
IPT
iPTP
iPDP
iPMP
CYP735A
tZTP
tZDP
tZMP
LOG
iPR
tZR
DZMP
DZR
iP
tZ
DZ
DMAPP + tRNA
tRNA-IPT
prenyl-tRNA
cZMP
cZR
cZ
Methionine
SAMS
SAM
ACS
ACC
ACO
Ethylene
GGDP
CPS
CDP
KO
ent-kaurenoic acid
KAO
GA13ox
GA12
GA53
GA20ox
GA9
GA20
GA3ox
GA4
GA1
Indole
Tryptophan
Tryptophan independent pathway
TAA
TAM
IAM
IPA
IAOx
IAAld
IAN
AMI1
YUC
AAO1
NIT1
IAA
Linolenic acid
LOX
13-HPOT
AOS
12,13-EOT
AOC
cis-(+)-OPDA
OPR3
OPC-8:0
(+)-7-iso-JA
Chorismate
ICS
Isochorismate
Phenylalanine
PAL
Cinnamate
Benzoate
Salicylic acid
trans-B-carotene
D27
9, cis-B-carotene
CCD7
9, cis-B-apo-10'-carotenal
CCD8
carlactone
MAX1
(+)-5-deoxystrigol
B
ABA
BR
CK
ET
GA
IAA
JA
SA
SL
ZEP
NCED
SDR (ABA2)
AAO
DET2
DWF4
CPD
IPT
LOG
tRNA-IPT
SAMS
ACS
ACO
CPS
KO
KAO
GA20ox3
GA3ox1
TAA
YUC
AMI1
AAO1
NIT1
LOX
AOS
AOC
OPR
ICS
PAL
D27
CCD7
CCD8
MAX1
Chloro-phyceae
Gonium pectorale
Volvox carteri
Chlamydomonas reinhardii
Ulva mutabilis
Trebouxiophyceae
Asterochloris sp.
Coccomyxa subellipsoidea
picochlorum sp.
Auxenochlorella protothecoides
Helicosporidium sp.
Chlorella sp. NC64A
Prasinophytes
Bathycoccus prasinos
Ostreococcus tauri
Ostreococcus leucimarinus
Ostreococcus sp. RCC809
Micromonas pusilla
Micromonas sp. RCC299
Streptophyta
Klebsormidium flaccidum
Arabidopsis thaliana
Oryza sativa
Physcomitrella patens

Figure S1. Phylogenetic position of the sequenced *Ulva mutabilis* strain. Maximum likelihood tree (log likelihood: -4386.98) of the *rbc*L gene (1,425 bp), based on a GTRGAMMA model with 25 rate categories and a single partition using RAxML v.8.2.4. Branch support results from a bootstrap analysis with 500 replicates. *Umbraulva olivascens* and *Ulvaria obscura* were used as outgroup.

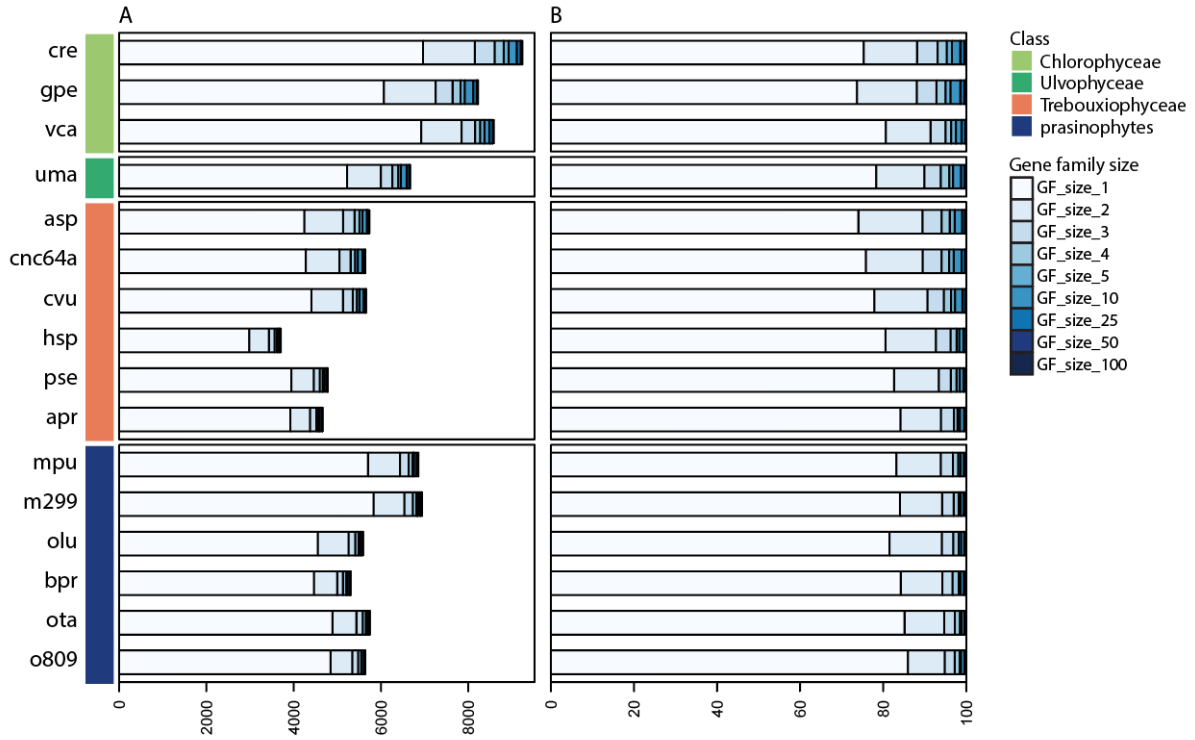


Figure S2. Distribution of gene family sizes in the green algal classes

A. actual numbers. B. percentages of total number of gene families. Abbreviations, see Table S8.

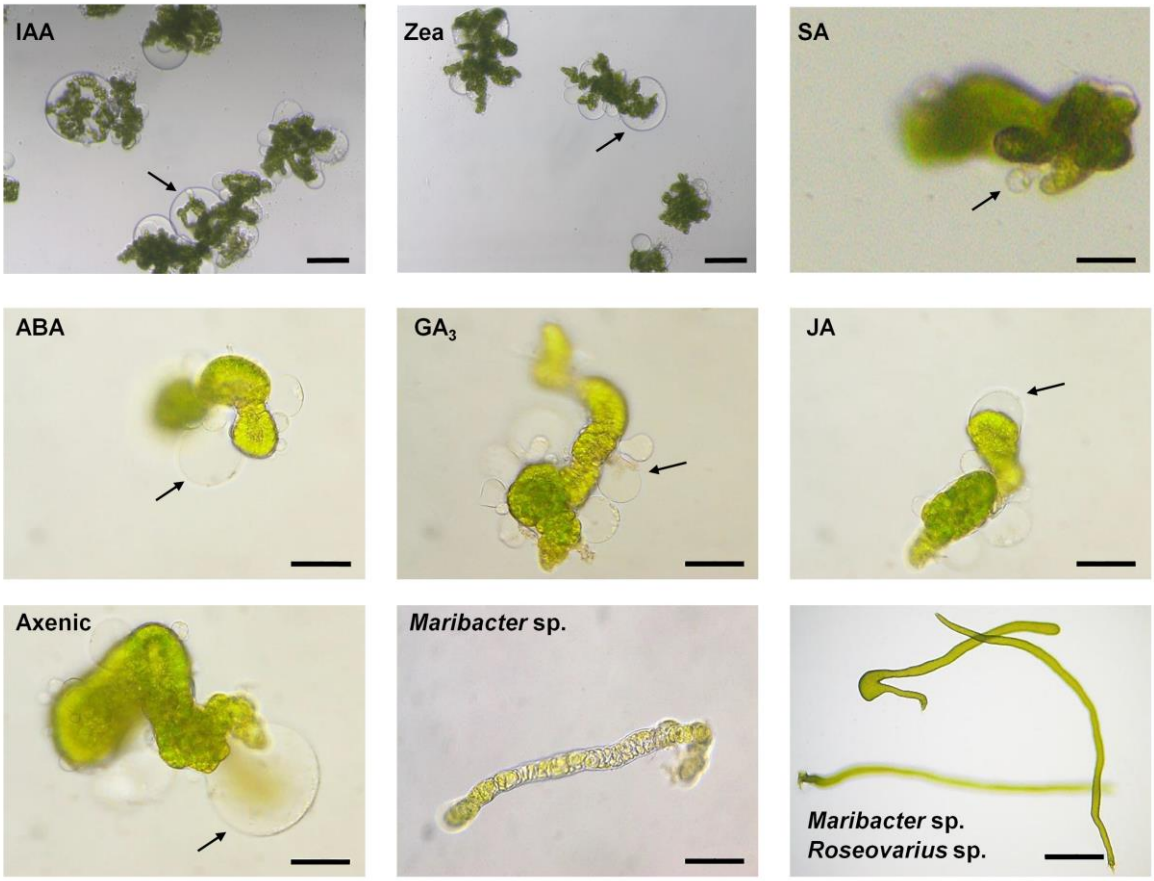
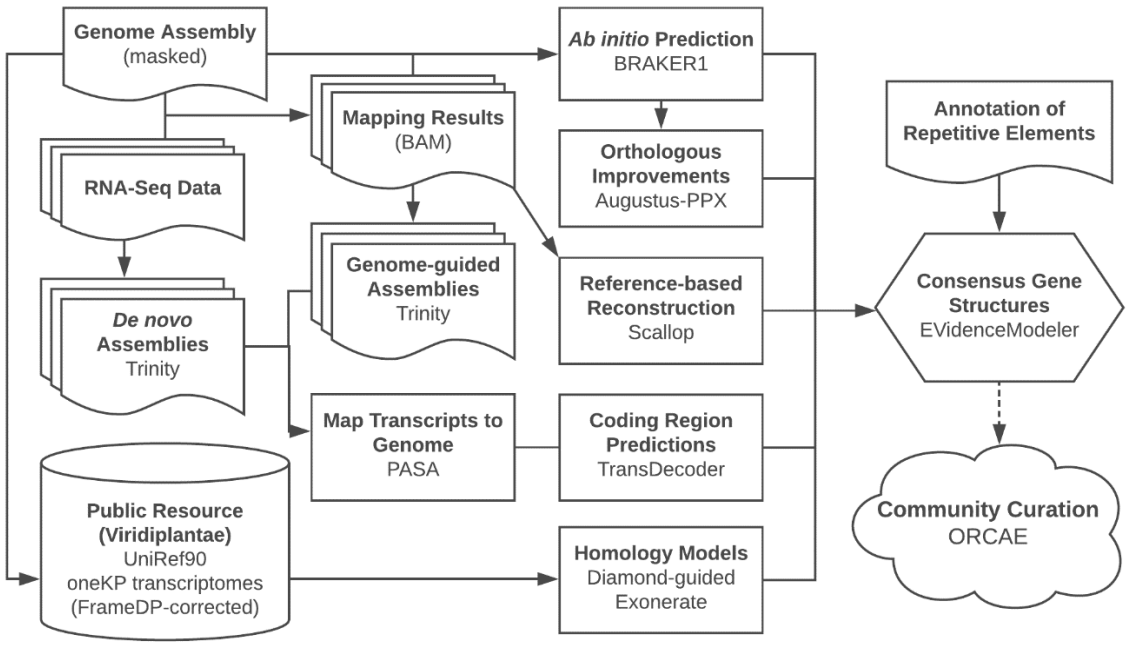


Figure S3. Morphogenetic bioassays of gametes inoculated with phytohormones. Representative algae of 2-weeks old germlings are shown. Protrusions of malformed cell walls are visible. Axenic gametes were inoculated with phytohormones (final concentration 10^{-9} mol/l). Phenotypes did not change at higher concentrations (10^{-6} mol/l). Top line: phytohormones IAA, Zea, SA, bars = 100 µm, middle line: phytohormones ABA, GA_3, JA, bars = 50 µm, bottom line: controls are shown under axenic conditions and inoculated with *Maribacter* sp. MS6 (bars = 50 µm). Growth and development of axenic cultures recovered upon inoculation with *Maribacter* sp. and *Roseovarius* sp. (bar = 500 µm). The combined application of all tested phytohormones did not recover the morphogenesis of *Ulva* either.

A



B

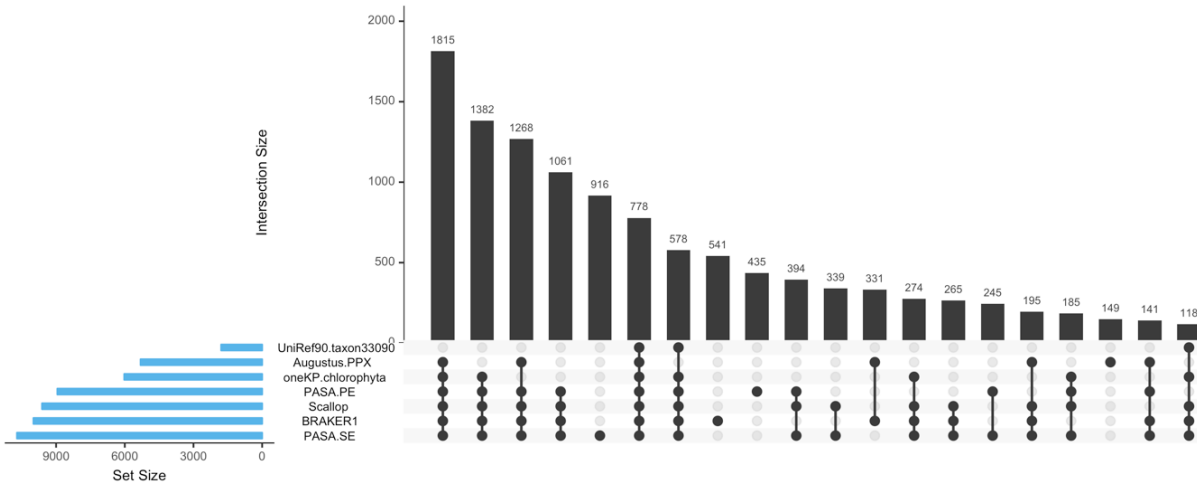


Figure S4. Gene prediction for the *Ulva mutabilis* genome.
A. Workflow of the gene prediction. B. Contributions of the different approaches to the final gene models. PASA.PE: predicted coding regions of PASA alignments using assembled RNA-Seq data from paired-end stranded data; PASA.SE: predicted coding regions of PASA alignments using assembled RNA-Seq data from single-end stranded data. The top 20 intersects are shown.

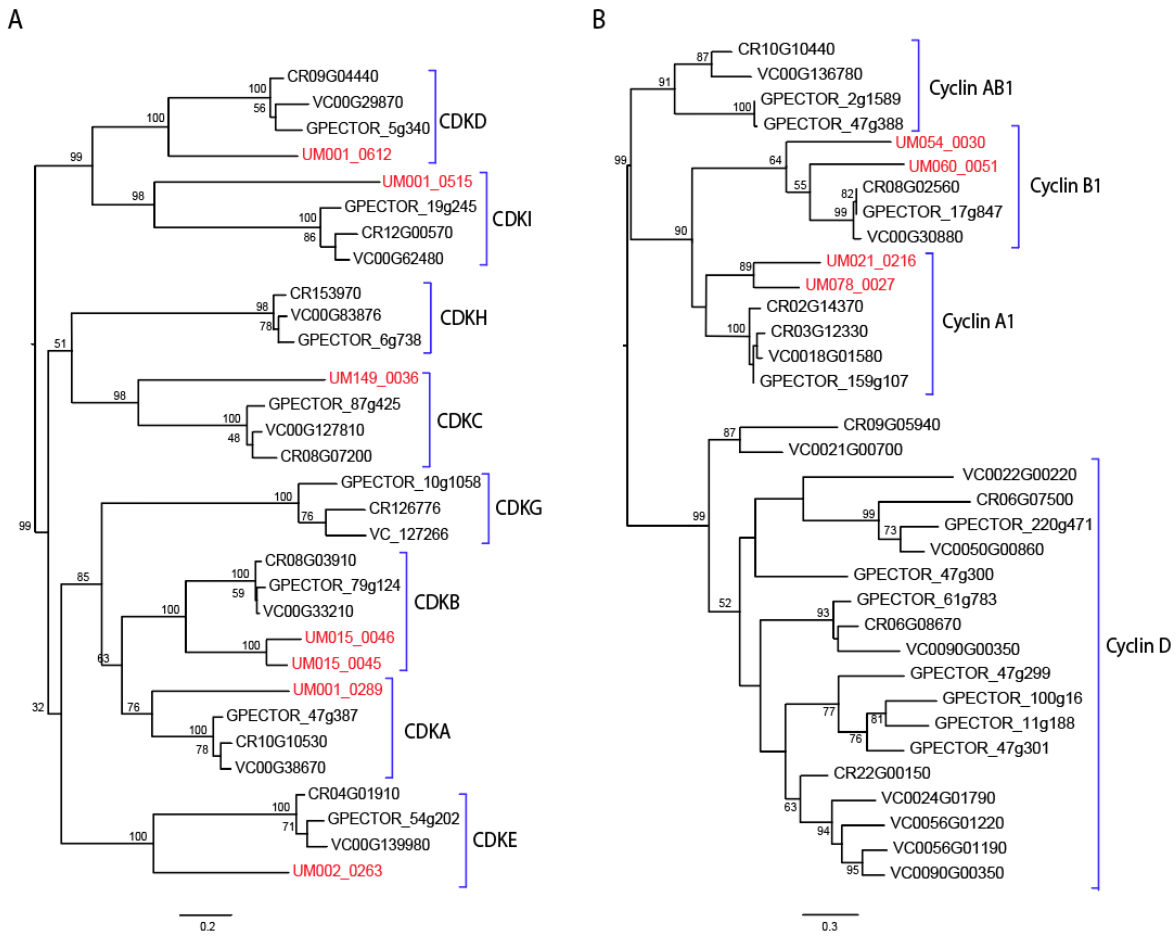


Figure S5. Phylogenetic analyses of cell cycle related genes in *Ulva mutabilis* and volvocine algae.

A. Cyclin Dependent Kinases (CDKs); B. Cyclins. Amino acid alignments were analysed using an automatically selected LG+GAMMA model under the Maximum Likelihood optimality criterion in RAxML. Bootstrap values result from fast bootstrapping algorithm implemented in RAxML. Assignment to different classes of CDK and cylin genes follows Hanschen et al. (2016). Abbreviations: *Chlamydomonas reinhardtii*, CR; Gonium pectoral, GPECTOR; *Volvox carteri*, VC.

Table S1. Assembly metrics of the *Ulva* genome.

Assembly metrics	**Scaffolds**	**Contigs**
Number of sequences (>= 1k bp)	318	381
Number of sequences (>= 5k bp)	315	378
Number of sequences (>= 10k bp)	312	372
Number of sequences (>= 25k bp)	276	322
Number of sequences (>= 50k bp)	243	274
Total length (>= 1k bp)	98,484,689	98,388,870
Total length (>= 5k bp)	98,477,802	98,381,983
Total length (>= 10k bp)	98,460,404	98,340,275
Total length (>= 25k bp)	97,848,084	97,481,650
Total length (>= 50k bp)	96,687,287	95,752,299
Largest sequence	3,623,364	2,868,306
Total length	98,484,689	98,388,870
GC (%)	57.2	57.2
N50	600,008	527,412
NG50	591,658	514,066
N75	340,001	289,826
NG75	330,723	267,228
L50	46	54
LG50	47	56
L75	100	117
LG75	103	122
Number of N's per 100 kbp	98	0

Table S2. Gene space completeness revealed by eukaryotic BUSCO and Core Gene Family analysis

Species	**Number of protein sequences**	**BUSCO**				**Core g**
		Single-copy Complete	Duplicated	Fragmented	Missing	Numbe
C. reinhardtii	17,741	285	6	7	5	1,812
C. subellipsoidea C169	9,629	265	5	14	19	1,787
D. salina	16,697	232	3	51	17	1,754
G. pectorale	16,290	243	4	39	17	1,779
M. pusilla ccmp1545	10,660	259	6	11	27	1,810
Micromonas RCC299	10,103	267	5	10	21	1,815
O. lucimarinus	7,796	230	26	11	36	1,815
U. mutabilis	12,924	264	2	14	23	1,717
V. carteri	14,247	281	6	6	10	1,795

Table S3. Summary of repetitive elements annotated in *Ulva mutabilis*.

Class	Family	Total bases	% masked
LTR	Copia	13,323,328	13.53
	Gypsy	1,607,922	1.63
	Pao	141,541	0.14
	LTR	102,691	0.10
	DIRS	72,745	0.07
	Ngaro	49,613	0.05
	ERVL	46,795	0.05
LINE	LINE	3,433,947	3.49
	CRE-II	2,753,584	2.80
	CRE-Cnl1	1,408,967	1.43
	R1	720,404	0.73
	I	483,199	0.49
	Penelope	269,147	0.27
	L1-Tx1	223,981	0.23
	RTE-X	75,371	0.08
	CR1	35,377	0.04
DNA	hAT-Ac	39,812	0.04
	Dada	30,836	0.03
	DNA	29,444	0.03
	hAT-Tag1	25,903	0.03
	hAT-Charlie	25,772	0.03
	CMC-EnSpm	7,698	0.01
SINE	Alu	90,465	0.09
RC	Helitron	46,685	0.05
Reported	RS280 (EU256376.1)	575,835	0.58
	RS360 (EU263359.1)	123,676	0.13
Unknown	Unknown	8,994,613	9.13

Table S4. Overview of phytohormones in axenic and non-axenic cultures (tripartite community with the bacterial strains *Roseovarius* sp. MS2 and *Maribacter* sp. MS6) of *Ulva mutabilis* (*wildtype* and *slender* strain) and the presence of proteins involved in their biosynthesis or signaling.

ND = not determined, LOQ = limit of quantification.

	LC-MS and GC/MS analyses (ng/g dry weight or nL/h.g FW)			**Pathways**		**Produced by bacteria MS2/MS6**
	Axenic (slender)	**Xenic (slender)**	**Xenic (wildtype)**	**Bio-synthesis**	**Intra-cellular signaling**	
Abscisic acid (ABA)	2.03 ± 0.18	2.57 ± 0.24	5.33 ± 0.27	Complete	No	No/No
Auxin (IAA)	48.0 ± 11.0	51.04 ± 5.79	58.38 ± 22.6	Complete	No	Yes/Yes
Ethylene (ET)	0.35 ± 0.22	0.3 ± 0.20	ND	Complete	No	ND
Gibberellin (GA_3)	< LOQ (traces)	< LOQ (traces)	No	Incomplete	No	Yes/Yes
Salicylic acid (SA)	124.1 ± 14.45	28.68 ± 4.97	41.67 ± 2.24	Complete	No	Yes/Yes
Brassino-steroids	No	No	No	No	No	ND
Cytokinins (Zeatin, iP)	No	No	No	Incomplete	No	Yes/Yes (iP)
Jasmonate (JA)	No	No	No	No	No	ND
Strigolactones	No	No	No	No	No	ND

Table S5. Horizontal gene transfer candidates in *Ulva mutabilis*

No.	Gene models	Annotation	Functional Category	Notes	Exon (nr)	GC (%)	Transcription (FPKM value)
1	UM005_0173	deacylase YbaK-like	Translation	Ensures translation fidelity	8	67.28	50.1
2	UM107_0008	50S ribosomal protein L18	Translation	Ribosome structural component. Mediates the attachment of the 5S RNA into the large ribosomal subunit in bacteria.	5	62.11	98.1
3	UM033_0031	ATP-dependent DNA helicase PcrA	DNA repair	DNA unwinding in bacteria. Belongs to UvrD/Rep helicase family.	8	59.42	73.4
	UM033_0032 (tandem dup.)				2	57.74	18.4
4	UM042_0001	DNA-3-methyladenine glycosylase	DNA repair	DNA alkylation repair	6	56.62	
	UM084_0052				5	52.05	24.6
5	UM013_0142	glucuronyl hydrolase	Carbohydrate metabolism	Glycosyl Hydrolase Family 88; Unsaturated glucuronyl hydrolase catalyzes the hydrolytic release of unsaturated glucuronic acids from oligosaccharides (EC:3.2.1.-) produced by the reactions of polysaccharide lyases.	7	64.01	17.7
	UM034_0033				5	66.36	32.5
6	UM035_0080	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase	Cell wall metabolism	MurF. Involved in cell wall formation. Catalyzes the final step in the synthesis of UDP-N-acetylmuramoyl-pentapeptide, the precursor of murein.	3	53.31	18.2
7	UM016_0037	Haem peroxidase	Response to stress	Reduces hydrogen peroxide	8	61.45	31.1
	UM001_0001				4	62.89	0.5
	UM001_0017				4	60.94	0
	UM001_0023				5	61.36	1.9
	UM001_0025				5	60.99	36.7
	UM001_0027				3	54.85	0.2

	UM001_0028				3	61.55	7
	UM001_0030				1	61.43	0
	UM001_0033				4	61.38	0
	UM017_0003				5	61.21	0.3
	UM017_0004				5	61.19	0
	UM017_0005				5	62.63	0
	UM018_0064				5	63.54	4.3
	UM018_0067				7	67.09	210.6
	UM024_0048				4	62.83	3.1
	UM024_0060				6	59.73	32.6
	UM049_0004				5	60.16	0.1
	UM049_0029				6	64.81	56.5
	UM061_0028				8	60.06	86.4
	UM077_0050				5	64.65	1
	UM104_0025				4	62.96	26.4
	UM104_0027				5	59.9	1.9
	UM104_0028				4	62.48	0.5
	UM104_0029				3	62.19	1
	UM104_0030				2	63.98	0.0
	UM154_0005				4	61.3	0.9
	UM160_0003				6	60.89	0.8
	UM169_0003				7	63.88	1.1
	UM180_0001				4	59.59	0.1
	UM180_0002				2	58.5	2.4
	UM180_0003				4	61.2	1.3
	UM180_0006				5	62.04	0.5
	UM180_0007				2	60.7	0.0
	UM199_0002				2	62.17	0.0
	UM290_0001				4	62.06	0.0

	UM290_0002				3	60.33	0.0
8	UM109_0045	cupin	other	The cupin superfamily of proteins is functionally extremely diverse. It was named on the basis of the conserved β-barrel fold	2	58.05	32.9
9	UM052_0037	D-lactate dehydrogenase	other	Catalyzes interconversion of pyruvate and lactate	4	61.19	1.8
10	UM031_0093	serine/threonine protein kinase	other	catalytic domain	7	66.53	47.5
	UM031_0094 (tandem dup.)				3	65.73	30.7
	UM031_0095 (tandem dup.)				6	67.22	13.3
	UM031_0097 (tandem dup.)				6	67.6	3.4
11	UM023_0015	multifunctional 2',3'-cyclic-nucleotide 2'-phosphodiesterase/5'-nucleotidase/3'-nucleotidase	other	nucleotide scavenging and transport	9	66.48	27.1
	UM035_0119				9	57.34	22.3
	UM042_0110				3	68.56	0.0
	UM042_0111				10	65.05	11.4
12	UM094_0024	serine/threonine protein kinase	other	catalytic domain	3	68.55	1.2
	UM001_0092				3	67.5	24.2
	UM001_0100				3	67.65	42.1
	UM006_0116				4	68.69	6.7
	UM011_0009				3	59.68	5.2
	UM025_0014				3	66.59	26.2
	UM173_0008				2	67.24	14.2
13	UM009_0111	DUF3179 domain-containing protein	unknown function		1	55.74	26

Table S6. DNA and RNA sequencing libraries.

Sequencing Platform	Source and Type	Strain	Read Number	Total bases (Gbp)
PacBio RSII (P6 chemistry)	DNA (>8 kb)	WT	770,598	6.91
Illumina MiSeq (2x250 bp paired-end)	DNA (insert size 350, 550 bp)	WT	16,398,654	2.92
Illumina HiSeq 2500 (2x125 bp paired-end)	DNA (insert size 2kb)	WT	170,149,048	17.75
Illumina HiSeq 2500 (2x125 bp paired-end)	DNA (insert size 5kb)	WT	166,928,626	17.50
Illumina HiSeq 2500 (2x125 bp paired-end)	DNA (insert size 10kb)	WT	126,821,228	13.13
Illumina HiSeq 2500 (2x125 bp paired-end)	RNA (FR-second strand)	WT, SL	397,961,040	44.67
NextSeq 550 (150 bp single-end)	RNA (reverse strand)	SL	164,378,893	22.56

WT: wild-type strain; SL: slender strain

Table S7. Species list and genome versions used for annotation and comparative genomics in a custom PLAZA platform.

Species	Source	PubmedID	Abbrev.
Aureococcus anophagefferens	JGI 1.0	21368207	aan
Asterochloris sp. Cgr/DA1pho v2.0	JGI 7.45.13	/	asp
Auxenochlorella protothecoides	Beijing Genomics Institute 1.0	25012212	apr
Arabidopsis thaliana	TAIR10	11130711	ath
Amborella trichopoda	Amborella v1	24357323	atr
Bathycoccus prasinos	Ghent University	22925495	bpr
Chondrus crispus	ENSEMBL release 28	23536846	ccr
Caenorhabditis elegans	ENSEMBL release 81	9851916	cel
Cyanidioschyzon merolae	Tokyo University	15071595	cme
Chlorella sp NC64A	JGI 1.0	20852019	cnc64a
Chlamydomonas reinhardtii	JGI 5.5 (Phytozome 10.2)	17932292	cre
Coccomyxa subellipsoidea C-169	JGI 2.0 (Phytozome 10.2)	22630137	csu
Drosophila melanogaster	ENSEMBL release 81	10731132	drm
Ectocarpus siliculosus	Ghent University	20520714	esi
Gonium pectorale	GenBank (LSYV01000000)	27102219	gpe
Homo sapiens	ENSEMBL release 81	11181995	hom
Helicosporidium sp.	Illinois University 1.0	24809511	hsp
Klebsormidium nitems	Tokyo Inst. Technology	24865297	kni
Micromonas pusilla strain CCMP1545	Ghent University	24273312	mpu
Micromonas sp RCC299	JGI 3.0	19359590	m299
Ostreococcus lucimarinus	JGI 2.0	17460045	olu
Ostreococcus sp RCC809	JGI 2.0	/	o809
Ostreococcus tauri	Ghent University v2.0	25494611	ota
Oryza sativa	MSU RGAP 7	16100779	osa
Physcomitrella patens	Phytozome 9.1 (v1.6)	18079367	ppa
Picochlorum sp. SENEW3 (SE3)	Rutgers University 1.0	24965277	pse
Phaeodactylum tricornutum	ASM15095v2	18923393	ptr
Saccharomyces cerevisiae strain S288C	ENSEMBL release 81	8849441	sac
Schizosaccharomyces pombe	ENSEMBL fungi release 28	11859360	scp
Thalassiosira pseudonana	JGI 3.0	15459382	tps
Ulva mutabilis	/	/	uma
Volvox carteri	JGI 2.0 (Phtyozome 10.2)	20616280	vca

Table S8. Target genes and associated primers for qPCR analysis of DMSP-related genes.

name	target gene identifier	primer sequence (5'->3')
p111f	UM008_0183 (Ubiquitin)	CCCTCGAAGTGGAGTCTTCTGAC
p112r	UM008_0183 (Ubiquitin)	AAGTGTGCGGCCATCCTCTA
p173f	UM010_0003 (PP2A 65 kDa regulatory subunit A)	GGCAACTGCAGGAGCAATTCT
p160r	UM010_0003 (PP2A 65 kDa regulatory subunit A)	CCTCAGAAGCAACCTCGACCAT
p169f	UM030_0039 (DMSP lyase)	TTCGACGACAAAGAGAAGATCGCA
p170r	UM030_0039 (DMSP lyase)	TAGCGGTCCTTCTCCAGGTC
p167f	UM021_0036 (DMSP lyase)	CAAGCCTGTGCTGCTCTCG
p171r	UM021_0036 (DMSP lyase)	GTGGCCGTTTGCCGTAAAGA
p174f	UM033_0146 (BCCT transporter)	CAACGCCAGGCTCCAAGAC
p175r	UM033_0146 (BCCT transporter)	CACCCATGCTTGCTGTGTGA
p180f	UM033_0147 (BCCT transporter)	GGGCGGTCAACATGTCGTT
p181r	UM033_0147 (BCCT transporter)	GGACTTCATCGTCATCGAGAACC
p184f	UM033_0150 (BCCT transporter)	AGGCCATCAATCTGACCGGATT
p185r	UM033_0150 (BCCT transporter)	AGACTTGAGTGTCATGGCAAAGC
p186f	UM033_0151 (BCCT transporter)	GGCCAGAGAAGTCAAATGCAGAG
p187r	UM033_0151 (BCCT transporter)	ACCACAAAGATGTTCTGAAGGCC
p190f	UM036_0102 (methyltransferase)	GCAATGGCTGAAGCATGCG
p191r	UM036_0102 (methyltransferase)	GCCCTTGCCTGAGCCTAAATC
p192f	UM052_0056 (methyltransferase)	ATCAACTTTGGCAACTCCCTGC
p193r	UM052_0056 (methyltransferase)	GTTGGGGCCAAGCAGAATG

92

Click here to access/download

Supplemental Videos and Spreadsheets

Suppl.Data1.xlsx