nature portfolio

Peer Review File

Genomic and transcriptomic basis of morphological and life cycle diversity in the prasinophyte alga *Pseudoscourfieldia marina*

Corresponding Author: Professor Claude Lemieux

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Crepeault et al. report the genome and transcriptome of two morphotypes, i.e., diploid and haploid stages, of a marine prasinophyte P. marina. They show that the genomes of CCMP1203 and UIO007 have genomic traits for haploid and diploid, respectively. The difference in gene expression patterns suggests potential genes for flagellar and scale. Based on these data, they reinforce the concept that CCMP1203 and UIO007 are in different cell cycle stages of a single species P. marina. Moreover, they found outlier chromosomes in this lineage and discussed their function in the viral resistance system. Overall, the manuscript was well-documented and easy to read; however, I have some concerns described below.

Major Points:
- You argued that CCMP1203 is the diploid stage of P. marina. Unfortunately, I could not find a description of the sex determination system in this species. Are there any potential genes, such as mid, for sex determination in the genome?
- If the morphotypes are in different stages in their life cycle, I feel a little strange about their variation in gene contents. How different is the evolutionary selective pressure of the putative flagellate or scale-related genes? If these genes have significantly weak selective pressure in CCMP1203, it can be a strain lacking the flagellate stage. Of course, I know that flagellate cells were observed in this strain during the initial isolation, but the observation was too old, and the flagellates have never been discovered since then to my best knowledge.
- In this manuscript, you discussed a close phylogenetic relationship between P. marina and C. primus, such as Fig. 2B. Do they form a sister group? This is inconsistent with the result of Yang et al. (2023). You should add Pradinoderma as an outgroup for the phylogenetic analysis.
-

Minor Points:
- L56. Yang et al. (2023) denied the phylogenetic position of Prasinodermatophyta reported in Li et al. (2020). Therefore, please remove the reference Li et al. here.
- L92. Table1. Please add the other available genomes of P. marina in Table 1.
- L99. “gene familier” is unclear. Did you mean that these genes are duplicated?
- L114. Please add k-mer analyses using raw reads. Are there two peaks for the CCMP1203 genome?
- L120. Triploid signals. Can you omit the possibility of the assembly error?
- L135-138. Are these genes shared between both types of cells?
- L148-149. You are better off showing a scatter plot of these transcription activities using such as RPKM.
- L159. How did you enrich the DEGs? Any statistical treatment?
- L165-171. This paragraph is redundant in the results section. Please shorten it, or move it to the introduction/discussion section.
- L172-173. You mean that the two strains had the same subset of genes for flagellar? Please clarify it.
- L185-191. This paragraph is redundant in the results section. Please shorten it, or move it to the introduction/discussion section.
- L241-243. I am confused...Why is the number of genes different if they are the same species with different cell cycle stages?
- L303. Have viral particles been observed using TEM? Did you find any viral sequence in your genome or transcriptome

reads?
- L308-309. Do these genomes (NIES-2893 and CCAP 190/2) encode the homologous genes on the outlier chromosomes in this study? Moreover please add “-” between NIES and 2893.
- L376. Is there any proof that this gene (25g105430) is a homolog of DUC1? P. marina possesses some usual genes for cryptochromes.
- L429. Please give us references for L1 medium and TL30 medium.
- L454. The sentence is unclear. Did you use the results of braker and augustus as input to maker? Why did you use both braker and augustus? Braker uses augustus inside, and thus, the addition of augustus simply can increase some errors.
- L464. Did you perform BUSCO analysis using protein mode?
- L475. How could you find telomere sequences? What is the telomere sequence?
- L506. Ultrafast bootstrapping in IQ-tree is a little bit unreliable. Please use non-parametric bootstrapping for small datasets like these.

Reviewer #2

(Remarks to the Author)
This study describes a genomic and transcriptomic comparative analysis of two morphologically different strains of the prasinophyte green alga Pseudoscourfieldia marina. The data presented strongly indicates that the two morphotypes correspond to the haploid and diploid life cycle phases of this species. Moreover, comparison of the two genomes has allowed the identification of genes with potential roles in determining morphological features specific to each strain (i.e. life cycle stage), such as the flagella and the extracellular scales of the haploid strain UIO007.

The authors also report several differences between the genomes of the two strains, including repeat content, chromosome rearrangements, aneuploidy and truncation of the gene DUC1. One possibility evoked is that these differences arose during the maintenance of the strains in culture. The authors suggest that differences at the DUC1 locus might contribute to vertical niche partitioning (line 380). This is an interesting hypothesis but they should make it clear that this would only be possible in situations where haploids have opted out of the sexual cycle because a functional DUC1 would be required again after transition to the diploid phase. This problem is alluded to but not specifically stated.

Analysis of these two genomes has proved to be a very effect means to obtain insights into the life cycle of this alga and the results obtained are very interesting but it is important not to ignore the fact that the life cycle of this alga cannot yet be completed in the laboratory and that this leaves a lot of questions unanswered. For example, how do the transitions from one phase to the other occur? Is the haploid strain capable of acting as a gamete or does it have to become fertile in some way before engaging in sexual fusion? How might the information gained here be used to develop a strategy to complete the life cycle in the laboratory? It would be good to have some discussion of these points in the manuscript, in particular to provide some context to the claim that Pseudoscourfieldia marina could be a powerful model to study life cycle evolution (line 420).

Finally, the authors mention life cycle regulators such as TALE homeodomain transcription factors but this aspect could be developed further. In particular, it would be relatively easy to carry out a search for some key life-cycle-related genes identified in other species such as TALE homeodomain transcription factors, mating type genes, core meiosis genes, etc. and report their presence or absence and their expression patterns across the two strains. An analysis of this type would add to the interest of the study.

Overall, this is a high-quality manuscript. The analyses have been carried out very carefully and the results are described and discussed in detail. It is clear that the authors have made a considerable effort to present and illustrate their results clearly. I also appreciated that the data generated have been made fully available through the public databases. It was a pleasure to review this manuscript, which stands out in the current context of increasingly poorly written papers.

Detailed remarks

line 393. The seminal work on TALE homeodomain proteins in Chlamydomonas (Lee et al. 2008) should also be cited here, and perhaps also Ectocarpus (Arun et al. 2019).

Supplementary table 1. Perhaps indicate the outlier chromosomes in this table with asterisks?

Reviewer #3

(Remarks to the Author)
In this manuscript, Lemieux et al. conducted comparative genomic and transcriptomic analyses of two P. marina morphotypes: CCMP1203 (coccoid, non-flagellated) and UIO007 (scaly, flagellated). The authors showed that CCMP1203 and UIO007 are diploid and haploid, respectively. Furthermore, comparative transcriptome analysis revealed upregulation of flagellar- and scale-related biosynthesis genes in UIO007, and identified a network of carbohydrate-active enzymes involved in scale formation.

This reviewer does not have any major criticisms regarding the content and recognizes the importance of this study; a small comment is provided below. The authors mainly discuss morphological changes during the life cycle in this paper. However, there are no photographs of these algae in the figures. To help readers understand more easily, it would be better to include

photographs or illustrations in the figures.

This study demonstrates that P. marina occupies an important evolutionary position for understanding the life cycle of Viridiplantae. I believe that establishing laboratory conditions enabling completion of its life cycle will be a key step toward further advancing this field. I wish the authors continued success in developing this promising line of research.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I would like to thank the authors for their excellent revision of this manuscript.
All my concerns have been addressed.
I recommend that this paper be published.

Reviewer #2

(Remarks to the Author)
I have reviewed the revised manuscript and the authors have comprehensively addressed all the points I raised. I therefore recommend that the manuscript be accepted for publication.

Reviewer #3

(Remarks to the Author)
I have reviewed the revised manuscript and have no further comments.
I am looking forward to seeing this published in Communications Biology.

RESPONSE TO REVIEWERS

We sincerely thank the three reviewers for their thoughtful and constructive comments, which have helped us strengthen our conclusions and substantially improve the quality of the manuscript. In revising the manuscript, we carefully considered and addressed all comments provided.

Reviewer #1

Crépeault et al. report the genome and transcriptome of two morphotypes, i.e., diploid and haploid stages, of a marine prasinophyte *P. marina*. They show that the genomes of CCMP1203 and UIO007 have genomic traits for haploid and diploid, respectively. The difference in gene expression patterns suggests potential genes for flagellar and scale. Based on these data, they reinforce the concept that CCMP1203 and UIO007 are in different cell cycle stages of a single species *P. marina*. Moreover, they found outlier chromosomes in this lineage and discussed their function in the viral resistance system. Overall, the manuscript was well-documented and easy to read; however, I have some concerns described below.

Major Points:

1- You argued that CCMP1203 is the diploid stage of *P. marina*. Unfortunately, I could not find a description of the sex determination system in this species. Are there any potential genes, such as mid, for sex determination in the genome?

RESPONSE (lines 294-300). This question was also asked by reviewer #2, but unfortunately, we are unable to provide a clear answer at this time. Our search for sex-determination genes yielded no clear homologs of *Chlamydomonas* mating-type specific genes such as *MID, MTD1, MTA1,* and *FUS1*. We identified six RWP-RK transcription factors that could potentially encode a MID-like gene, which in *Chlamydomonas* plays a central role in mating-type determination by activating *minus*-specific genes and repressing *plus*-specific genes. However, the RWP-RK domains of the *P. marina* proteins were too divergent from the conserved MID domain of volvocine green algae to allow unambiguous identification of a true MID homolog. We have added these observations to the Results section and discussed briefly the implications (see also our response to point 3 of reviewer 2). The absence of known sex-determination genes in *P. marina* does not preclude the existence of a sexual cycle, but may suggest that sex determination in this alga involves lineage-specific or yet unidentified regulators.

2- If the morphotypes are in different stages in their life cycle, I feel a little strange about their variation in gene contents. How different is the evolutionary selective pressure of the putative flagellate or scale-related genes? If these genes have significantly weak selective pressure in CCMP1203, it can be a strain lacking the flagellate stage. Of course, I know that flagellate cells were observed in this strain during the initial isolation, but the observation was too old, and the flagellates have never been discovered since then to my best knowledge.

RESPONSE (lines 183-187; lines 314-317; Supplementary Tables 3 to 7). We thank the reviewer for this insightful suggestion. Upon revisiting our manuscript, we realized that there was no mention of the levels of amino acid conservation for flagellar and scale-related proteins between the two strains. To address this point, we have added a column showing pairwise sequence identities between CCMP1203 and UIO007 proteins in Supplementary Tables 3 to 7. These values reveal an extremely high degree of conservation: for example, among the 274 proteins involved in flagellar assembly, 85 show 100% sequence identity, with an average identity of 99.24% across the entire set. The results thus strongly support functional retention of these genes in both strains, rather than pseudogenization or relaxed selection.

We also considered performing dN/dS analyses to formally assess selective pressures; however, given that we only have two closely related strains (UIO007 and CCMP1203), pairwise dN/dS comparisons would not provide robust or interpretable results. A rigorous evaluation of evolutionary pressures on these gene families will require broader comparative analyses across additional *P. marina* isolates or related taxa, which we propose as an important direction for future research.

We have expanded the Discussion to emphasize that morphological divergence between strains is primarily driven by differential expression of flagellar and scale-related genes rather than pseudogenization or relaxed selection. Furthermore, the complete set of meiosis-related genes we uncovered in both genomes during our revision (see response to point 3 of reviewer 2) argues against the hypothesis that the diploid strain has permanently lost the ability to transition to a flagellated stage. As observed in other systems, switching between vegetative phases likely requires specific environmental cues that remain unidentified.

3- In this manuscript, you discussed a close phylogenetic relationship between *P. marina* and *C. primus*, such as Fig. 2B. Do they form a sister group? This is inconsistent with the result of Yang et al. (2023). You should add Prasinoderma as an outgroup for the phylogenetic analysis.

RESPONSE (lines 104-109; lines 554-556; Fig. 2). We greatly appreciated this comment. Yes, the phylogenomic analysis presented in our original manuscript did place *P. marina* as sister to the two *Chloropicon* species, which is inconsistent with the phylogenetic position reported by Yang et al. (2023). Following the reviewer's suggestion, we re-ran the phylogenomic analysis using *Prasinoderma coloniale* as the outgroup instead of the streptophyte *Klebsormidium flaccidum*. We also adopted the LG+C20+F+G4 evolutionary model employed by Yang et al. (2023) to ensure methodological consistency. In the resulting tree (revised Fig. 2b), the *P. marina* strains form an independent branch that diverges prior to the *Chloropicon* lineage. This result, which is now in full agreement with Yang et al. (2023), strengthens the phylogenetic interpretation presented in the manuscript.

Minor Points:

4- L56. Yang et al. (2023) denied the phylogenetic position of Prasinodermatophyta reported in Li et al. (2020). Therefore, please remove the reference Li et al. here.

RESPONSE (lines 55-56). As requested, we have removed the reference to Li et al. (2020). The phylogenetic position of Prasinodermophyta has been debated, with Li et al. (2020) reporting that it represents a third phylum of Viridiplantae, in conflict with earlier studies suggesting that it represents the most basal lineage of Chlorophyta. Yang et al. (2023) addressed this issue through comprehensive phylogenomic analyses using multiple datasets and methods, concluding that Prasinodermophyta is sister to all other chlorophyte lineages. Given this robust evidence, we consider the matter resolved and have revised the manuscript accordingly. The text now reflects that the Viridiplantae comprises two major lineages: Streptophyta and Chlorophyta.

5- L92. Table1. Please add the other available genomes of *P. marina* in Table 1.

RESPONSE (lines 96-99, Table 1). As requested, we have modified Table 1 and the text of the Results to include the genomes of the *P. marina* strains NIES-2893 and CCAP 190/2. Because the genes of strain CCAP 190/2 are not annotated, some genome metrics could not be provided for this strain.

6- L99. "gene families" is unclear. Did you mean that these genes are duplicated?

RESPONSE (lines 101-103). In the revised manuscript, we have replaced the ambiguous term "gene families" by the more precise phrase "orthogroups containing multiple paralogs". This change clarifies that we are referring to groups of homologous genes that include duplicated copies within the same genome.

7- L114. Please add k-mer analyses using raw reads. Are there two peaks for the CCMP1203 genome?

RESPONSE (lines 121-124; lines 521-523; Supplementary Fig. 3). We thank the reviewer for this valuable suggestion. We performed k-mer analyses using raw reads for both genomes. Specifically, we used Jellyfish v2.3.1 to count k-mers and generate count histograms, which were then plotted with GenomeScope 2.0. As expected, the GenomeScope profile obtained for CCMP1203 displayed two distinct peaks, consistent with diploidy, whereas that obtained for UIO007 showed a single peak, consistent with haploidy. To present these results in the revised manuscript, we included additional sentences in the Methods and Results sections, added the ploidy smudge plots and the underlying data generated by GenomeScope to the revised Supplementary Fig. 3, and also deposited on GitHub the command lines (as a shell script named kmer_ploidy_counts.sh) used for the analysis.

8- L120. Triploid signals. Can you omit the possibility of the assembly error?

RESPONSE (lines 125-133; Supplementary Fig. 4). We addressed this important point during our revision. While we cannot completely rule out the possibility of an assembly error, several lines of evidence strongly suggest that the region of UIO007 chromosome 31 showing triploid signals is correctly assembled. First, UIO007 chromosome 31 is highly syntenic and colinear along its entire length with the corresponding chromosome XXVI of CCMP1203, which does not

exhibit abnormal ploidy signals. Second, this chromosome also shows strong synteny with homologous chromosomes in the CCAP 190/2 and NIES-2893 genome assemblies (see alignments provided in Supplementary Fig. 4).

With regards to this point, the sentences we modified in the manuscript now read as follows: "Illumina read-depth analysis revealed aneuploidy in both strains. In CCMP1203, chromosome 32 occurs as a single copy, whereas chromosome 31 exhibits mixed diploid and triploid signals, possibly reflecting aneuploid variation within the cell population (Fig. 1). Although assembly errors cannot be entirely ruled out, synteny analysis supports assembly accuracy: UIO007 chromosome 31 is fully colinear with its homolog in CCMP1203, which shows no abnormal ploidy signal, and is highly syntenic with homologous chromosomes in the CCAP 190/2 and NIES-2893 assemblies (Supplementary Fig. 4). These syntenic relationships strongly suggest that the observed triploid signal represents genuine biological variation rather than an assembly artifact."

9- L135-138. Are these genes shared between both types of cells?

RESPONSE (lines 146-153; Supplementary Table 3). To clarify this point, we have revised the paragraph to indicate more clearly that the genes located on the outlier chromosomes represent a combination of conserved and strain-specific components. Furthermore, we have removed the old Supplementary Fig. 3 and replaced it with the new Supplementary Table 3 that lists all proteins with predicted functions or functional domains that are not retrotransposon-related, along with their conservation status between the two strains.

10- L148-149. You are better off showing a scatter plot of these transcription activities using such as RPKM.

RESPONSE (Fig. 4). In response to this excellent suggestion, we have replaced the old gene expression profiles of the outlier chromosomes by scatter plots representing transcript levels expressed as RPKM (Reads Per Kb per Million mapped reads). The plots clearly illustrate that genes on UIO007 chromosome 16 exhibit higher expression levels compared to those on CCMP1203 chromosome 32, reinforcing the transcriptional asymmetry between these regions.

11- L159. How did you enrich the DEGs? Any statistical treatment?

RESPONSE (lines 170-178; Supplementary Fig. 5 and its legend). Thank you for raising this point. We did not perform a formal enrichment test because the analysis was exploratory, intended to provide functional context and a transition between genomic and transcriptomic results. Our approach focused on identifying KEGG functional categories contributing the most to three expression groups defined by $\log_2$FC intervals ($\geq$2.00, $\leq$–2.00, and intermediate). Fig. 4b therefore shows the proportion of genes from each KEGG category within these groups, calculated relative to the total number of genes in each group. This representation highlights functional categories most likely associated with morphological differences.

We revised the entire paragraph of this Results section to clarify our approach, which now reads as follows: "Among the 8,647 single-copy orthologs shared by CCMP1203 and UIO007, 4,385 were differentially expressed (DEGs), with 3,369 upregulated in UIO007 and 1,016 in CCMP1203 (Supplementary Fig. 5a). To provide functional context, we grouped DEGs into three expression groups based on log_2FC intervals (≥2.00, ≤-2.00, and intermediate) and classified them into KEGG functional categories. Supplementary Fig. 5b shows the proportion of genes from each KEGG category within these groups, calculated relative to the total number of genes in each group. Notably, cilium proteins contributed most to the group of genes upregulated in the flagellated UIO007 strain, alongside cytoskeletal proteins. In contrast, CCMP1203 showed higher contributions from transporter and exosome protein families."

12- L165-171. This paragraph is redundant in the results section. Please shorten it, or move it to the introduction/discussion section.

RESPONSE (lines 359-361). We agree with the reviewer that this paragraph is not essential for the presentation of the results. We moved a shorter version of this paragraph to the Discussion section, where it provides better context for interpreting our findings.

13- L172-173. You mean that the two strains had the same subset of genes for flagellar? Please clarify it.

RESPONSE (lines 180-181). Thank you for pointing this out. We confirm that the two strains share the same subset of genes. To clarify, we revised the sentence as follows: "We identified a shared set of 274 genes involved in flagellar assembly in both CCMP1203 and UIO007 (Fig. 5a, Supplementary Table 4)."

14- L185-191. This paragraph is redundant in the results section. Please shorten it, or move it to the introduction/discussion section.

RESPONSE (lines 196-200). We shortened this paragraph and fused it with the next one. The revised sentences now read as follows: "Prasinophyte scales are synthesized in the Golgi apparatus and contain the 2-keto sugar 3-deoxy-D-manno-octulosonate (Kdo) as a major carbohydrate component, yet the enzymatic machinery responsible for their biosynthesis remains poorly characterized. The three genes involved in the CMP-Kdo biosynthesis pathway in land plants *(kdsA, kdsB*, and *kdsD)* are conserved in *P. marina*, and all are upregulated in UIO007 (Fig. 6, Supplementary Table 5) …"

15- L241-243. I am confused...Why is the number of genes different if they are the same species with different cell cycle stages?

RESPONSE (lines 250-252). Thank you for raising this point. The difference in gene numbers reflects strain-specific genes rather than cell cycle stage. The two additional GT23 genes in UIO007 are located on the outlier chromosome (chromosome 16), which contains 264 genes unique to this strain. It is important to note that the two *P. marina* strains characterized in our study, despite belonging to the same species, were collected from distant geographical locations and likely represent distinct populations. This geographic and population separation may have

facilitated the accumulation of structural variations, particularly in rapidly evolving genomic regions carrying non-essential genes.

16- L303. Have viral particles been observed using TEM? Did you find any viral sequence in your genome or transcriptome reads?

RESPONSE (lines 345-347). To our knowledge, viral particles have not been previously documented in TEM studies of *P. marina* cells. We did not conduct any TEM observations in the course of our study and no contigs of viral origin were identified in our CCMP1203 and UIO007 genome assemblies. This why we say that identifying and characterizing viruses that interact with *P. marina* will be critical for testing the hypothesis that outlier chromosomes could be involved in viral defense.

17- L308-309. Do these genomes (NIES-2893 and CCAP 190/2) encode the homologous genes on the outlier chromosomes in this study? Moreover, please add "-" between NIES and 2893.

RESPONSE (lines 164-168). We have corrected the strain name to include the hyphen (NIES-2893). To address the question, we searched the NIES-2893 genome for the genes identified on the outlier chromosomes of CCMP1203 and UIO007. As indicated in a new paragraph of Results, reciprocal BLASTP searches revealed no orthologs of these genes in the NIES-2893 genome. For CCAP 190/2, we could not perform a similar gene search because this genome has not been annotated yet by the Wellcome Sanger Institute.

18- L376. Is there any proof that this gene (25g105430) is a homolog of DUC1? *P. marina* possesses some usual genes for cryptochromes.

RESPONSE (lines 420-433; Supplementary Fig. 8). To address this point, we have added a new figure (Supplementary Fig. 8) comparing the domain architecture of DUC1 in NIES-2893, CCMP1203, and UIO007. All domains identified in NIES-2893 are conserved in CCMP1203, confirming that the latter strain possesses a true homolog of DUC1 (99.05% amino acid identity and 94% query coverage between the two strains). In contrast, UIO007 retains only a portion of the cryptochrome domain (98.15% amino acid identity and 29% query coverage when compared to NIES-2893 DUCI), consistent with a truncated version of the gene.

In the course of our revision, we found that this observation for UIO007 contrasts with previous findings by Makita et al. (2021), who reported an intact DUCI in the flagellated *P. marina* NIES-1419 strain (East China Sea). Because this discrepancy is incompatible with the hypothesis of systematic ecological divergence between haploid and diploid strains, we now believe that the observed truncation of DUCI in UIO007 most likely reflects a gene-loss event, perhaps resulting from its long-term maintenance (over 40 years) in culture collections (see also our response to point 1 of reviewer 2).

19- L429. Please give us references for L1 medium and TL30 medium.

RESPONSE (lines 472-475). As requested, we have added references to the NORCCA collection website for the compositions of both L1 (https://norcca.scrol.net/node/3954) and TL30 (https://norcca.scrol.net/node/3956) media.

20- L454. The sentence is unclear. Did you use the results of braker and augustus as input to make? Why did you use both braker and augustus? Braker uses augustus inside, and thus, the addition of augustus simply can increase some errors.

RESPONSE (lines 500-505). Thank you for pointing this out. We agree that the original sentence was unclear. To clarify, we made the following changes in the revised manuscript:
"Preliminary gene annotations were generated using Augustus 3.3.353 within the MAKER 2.31.10 pipeline to assess assembly quality prior to the availability of RNA-seq data. Once RNA-seq data became available, reads were normalized with Trinity 2.12.0 and mapped to the genome using HiSAT 2.1.0, enabling genome annotation to be performed with BRAKER 2.1.4. Both sets of predictions were subsequently loaded into Web-Apollo v2.5.0 for consistency checks and manual curation of gene models."

21- L464. Did you perform BUSCO analysis using protein mode?

RESPONSE (lines 512-514). Yes, we performed the BUSCO analysis in protein mode. This detail has been added to the revised manuscript to clarify our approach.

22- L475. How could you find telomere sequences? What is the telomere sequence?

RESPONSE (lines 91-93; lines 525-527). We modified a sentence of the Methods to clearly indicate that we used the custom Perl script check_for_telomeres.pl (v0.4b, available on GitHub) to identify telomeric sequences. Briefly, the script analyzes the first and last 1,000 bases of each contig, decomposes them into 5–8 nt k-mers to detect abundant motifs, and searches for a user-defined repeat pattern (default: TTAGG) occurring at least 10 times.

Following the reviewer's comment, we realized that the telomere repeat unit was missing from the original manuscript. Therefore, to report this sequence, we slightly modified one sentence of the Results that now reads as follows: "In the CCMP1203 strain, all chromosomes were fully sequenced from telomere to telomere (repeat unit: [TT]TTAGGG), whereas in the UIO007 strain, five chromosomes remain incomplete, each lacking a telomeric region at one end (Supplementary Table 1)."

23- L506. Ultrafast bootstrapping in IQ-tree is a little bit unreliable. Please use non-parametric bootstrapping for small datasets like these.

RESPONSE (lines 554-555; Fig. 2b). We agree that ultrafast bootstrapping can overestimate branch support, particularly for small datasets. As suggested, we re-ran the phylogenetic analysis using non-parametric bootstrapping. The updated tree (Fig. 2b) shows that all branches are highly supported, except for the branch leading to the *Pseudoscourfieldia* clade, which is supported by 82% bootstrap values.

Reviewer #2

This study describes a genomic and transcriptomic comparative analysis of two morphologically different strains of the prasinophyte green alga *Pseudoscourfieldia marina*. The data presented strongly indicates that the two morphotypes correspond to the haploid and diploid life cycle phases of this species. Moreover, comparison of the two genomes has allowed the identification of genes with potential roles in determining morphological features specific to each strain (i.e. life cycle stage), such as the flagella and the extracellular scales of the haploid strain UIO007.

The authors also report several differences between the genomes of the two strains, including repeat content, chromosome rearrangements, aneuploidy and truncation of the gene DUC1. One possibility evoked is that these differences arose during the maintenance of the strains in culture.

1-The authors suggest that differences at the DUC1 locus might contribute to vertical niche partitioning (line 380). This is an interesting hypothesis but they should make it clear that this would only be possible in situations where haploids have opted out of the sexual cycle because a functional DUC1 would be required again after transition to the diploid phase. This problem is alluded to but not specifically stated.

RESPONSE (lines 427-437). In responding to this comment, we revisited the literature and found that Makita et al. (2021) had previously examined the presence of DUC1 by PCR analysis in another flagellated strain of *P. marina* (NIES-1419). This detail initially escaped our attention because it was only reported in the supplementary material. Their finding of an intact DUC1 gene in NIES-1419 is incompatible with our earlier hypothesis of vertical niche partitioning between diploid and haploid strains. In light of this evidence, we now favor the interpretation that the truncation of DUC1 in UIO007 is most likely an artefact resulting from its long-term maintenance (over 40 years) in culture collections rather than a natural ecological adaptation.

We included additional sentences in the Discussion section to reflect this interpretation, which reads as follows: “This initially suggested a possible role in vertical niche partitioning, with diploid strains such as CCMP1203 potentially adapted to deeper or more variable light environments than their haploid counterparts. However, Makita et al. reported an intact DUC1 in another flagellated strain (NIES-1419), contradicting the hypothesis of systematic ecological divergence between haploid and diploid strains. Taken together, these findings suggest that the truncation observed in UIO007 most likely reflects a gene-loss event, perhaps resulting from its prolonged maintenance in culture collections since 1983. In this context, DUC1 function may have been compensated by one of the two cryptochrome genes that *P. marina* UIO007 shares with plants and other chlorophytes (RI054_26g108900 and RI054_26g109230). Comparative genomic surveys across a broader range of isolates will be critical to assess whether the lack of a functional DUC1 is widespread among natural *P. marina* populations.”

2-Analysis of these two genomes has proved to be a very effect means to obtain insights into the life cycle of this alga and the results obtained are very interesting but it is important not to ignore the fact that the life cycle of this alga cannot yet be completed in the laboratory and that this leaves a lot of questions unanswered. For example, how do the transitions from one phase to the

other occur? Is the haploid strain capable of acting as a gamete or does it have to become fertile in some way before engaging in sexual fusion? How might the information gained here be used to develop a strategy to complete the life cycle in the laboratory? It would be good to have some discussion of these points in the manuscript, in particular to provide some context to the claim that *Pseudoscourfieldia marina* could be a powerful model to study life cycle evolution (line 420).

RESPONSE (lines 438-449). We thank the reviewer for these very insightful comments. We have added the following paragraph in the Discussion to address these questions:

"While our genomic and transcriptomic data support a heteromorphic haplodiplontic life cycle in *P. marina*, the complete life cycle has not yet been completed under laboratory conditions. This leaves key questions unresolved: How do transitions between haploid and diploid phases occur? Can the haploid strain act directly as a gamete, or must it undergo further differentiation to become sexually competent? The presence of meiosis-specific genes and conserved transcriptional regulators suggests that the molecular machinery for sexual reproduction is intact, but its activation may depend on environmental cues or culture conditions not yet replicated in culture. Establishing conditions that trigger phase transitions, such as manipulating light regimes, or nutrient availability, will be critical to validate *P. marina* as a model for studying life cycle evolution in prasinophytes. Resolving these questions will not only clarify the biology of *P. marina* but also provide broader insights into the evolution of sexual reproduction and ploidy regulation in early-diverging green algae."

3-Finally, the authors mention life cycle regulators such as TALE homeodomain transcription factors but this aspect could be developed further. In particular, it would be relatively easy to carry out a search for some key life-cycle-related genes identified in other species such as TALE homeodomain transcription factors, mating type genes, core meio0sis genes, etc. and report their presence or absence and their expression patterns across the two strains. An analysis of this type would add to the interest of the study.

RESPONSE (lines 267-300; lines 398-411; Supplementary Table 7). We thank the reviewer for proposing a search for key life-cycle-related genes in *Pseudoscourfieldia*. These analyses are a great addition to our study.

We investigated whether *P. marina* possesses the genetic toolkit required for meiosis and syngamy. The results are reported in a new section of the Results entitled "Genomic signatures of meiosis and life cycle regulation in *P. marina*", and are summarized in an additional Table (Supplementary Table 7). Of the 33 genes in the established meiosis toolkit, 29 (including 10 meiosis-specific genes) were identified in both strains, with most showing higher expression in the diploid strain. We also detected two components of the conserved syngamy triad (HAP2/GCS1 and KAR5/GEX1), with expression patterns consistent with roles in gamete membrane and nuclear fusion, while FUS1/GEX2 was not found. Additionally, homologs of *Gsm1* and *Gsp1*, encoding regulators of the haploid-to-diploid transition in *Chlamydomonas*, were present and upregulated in the haploid strain, suggesting a regulatory mechanism distinct from that in volvocine algae. Finally, although no canonical mating-type genes were detected, we identified nine RWP-RK transcription factors with MID-like domains, albeit too divergent for

definitive assignment. These findings indicate that *P. marina* retains key components of the molecular machinery for sexual reproduction, despite the absence of direct evidence for syngamy.

Furthermore, the following sentences were added to the Discussion: "The identification of an almost complete set of meiosis-specific genes and their higher expression in CCMP1203 suggest that meiotic processes may occur during the diploid phase. Conversely, the upregulation of *Hap2/Gcs1* in the haploid UIO007 strain and the slightly higher expression of *Kar5/Gex1* in the diploid CCMP1203 strain indicate potential roles in membrane and nuclear fusions during syngamy. Intriguingly, homologs of the TALE-class transcription factors *Gsm1* and *Gsp1*, which regulate the haploid-to-diploid transition in *Chlamydomonas*, were both present and upregulated in UIO007. This pattern resembles the co-expression of the TALE homeodomain life-cycle regulators OUROBOROS or SAMSARA in both gametes of the brown alga *Ectocarpus* prior to diploid sporophyte formation. Co-expression of *Gsm1* and *Gsp1* in haploid cells suggests that activation of the heterodimeric transcription factor in diploid strains may involve mechanisms distinct from those described in *Chlamydomonas*. The absence of canonical mating-type genes, despite the presence of divergent RWP-RK transcription factors, further points to lineage-specific strategies for sex determination in *P. marina*."

Overall, this is a high-quality manuscript. The analyses have been carried out very carefully and the results are described and discussed in detail. It is clear that the authors have made a considerable effort to present and illustrate their results clearly. I also appreciated that the data generated have been made fully available through the public databases. It was a pleasure to review this manuscript, which stands out in the current context of increasingly poorly written papers.

We sincerely thank the reviewer for this positive and encouraging feedback. Your comments are highly motivating and reinforce our commitment to maintaining high standards in our work.

Detailed remarks

4-line 393. The seminal work on TALE homeodomain proteins in *Chlamydomonas* (Lee et al. 2008) should also be cited here, and perhaps also *Ectocarpus* (Arun et al. 2019).

RESPONSE (lines 290-292, lines 404-407). Thank you for this suggestion. We have cited both references (Lee et al., 2008; Arun et al., 2019) in the revised manuscript at appropriate locations.

5-Supplementary table 1. Perhaps indicate the outlier chromosomes in this table with asterisks?

RESPONSE (Supplementary Table 1). We have highlighted the outlier chromosomes with asterisks in the revised Supplementary Table 1 to make them easily identifiable.

Reviewer #3

In this manuscript, Lemieux et al. conducted comparative genomic and transcriptomic analyses of two *P. marina* morphotypes: CCMP1203 (coccoid, non-flagellated) and UIO007 (scaly, flagellated). The authors showed that CCMP1203 and UIO007 are diploid and haploid, respectively. Furthermore, comparative transcriptome analysis revealed upregulation of flagellar- and scale-related biosynthesis genes in UIO007, and identified a network of carbohydrate-active enzymes involved in scale formation.

1-This reviewer does not have any major criticisms regarding the content and recognizes the importance of this study; a small comment is provided below. The authors mainly discuss morphological changes during the life cycle in this paper. However, there are no photographs of these algae in the figures. To help readers understand more easily, it would be better to include photographs or illustrations in the figures.

RESPONSE (Supplementary Fig. 1). Thank you for this valuable suggestion. We have added a new figure presenting SEM micrographs of the CCMP1203 and UIO007 strains (Supplementary Fig. 1). While we considered including these images in the main figures, space limitations prevented us from displaying them at sufficient resolution, so we opted to provide them as supplementary material.

This study demonstrates that *P. marina* occupies an important evolutionary position for understanding the life cycle of Viridiplantae. I believe that establishing laboratory conditions enabling completion of its life cycle will be a key step toward further advancing this field. I wish the authors continued success in developing this promising line of research.