Dear Editor,

We thank you and the reviewers for their evaluation of our submitted work. We are herewith resubmitting a carefully revised version of the manuscript including textual changes, additional analyses and responses to the comments raised by the reviewers and yourself. We sincerely hope that, with our comprehensive modifications, this genome note will be acceptable for publication in PCI Genomics.

Hanna Rovenich on behalf of all authors.

Editor

Dear authors,
Thank you very much for submitting your study to PCI Genomics. Your study has been seen by four Reviewers, who provided thorough comments that I believe could help further improve the manuscript. Given that this is a genome note, I do not think it needs to evolve into a comparative genomics paper, but I would appreciate a bit more context regarding the interest of sequencing this genome and the availability of further *Coccomyxa* genomes in NCBI/ENA.
>> We thank the editor for their response. The main interest to sequence the genome of *Coccomyxa viridis* is their widespread existence as symbiotic phototrophs, especially their general occurrence in lichens, which has recently been confirmed (https://www.nature.com/articles/s41598-023-48637-w). Despite the fact that lichens have been studied for over a century, little is known about the underlying mechanisms that drive these symbioses. The availability of high-quality genomes, such as the one generated here, will facilitate the study of these fascinating systems. This is now explained in the manuscript text.
The only other high-quality genome of *C. viridis* has been made available following the upload of the first version of this manuscript to BioRxiv and is now referred to in the text. In addition to the genome of *C. subellipsoidea*, which was included in this manuscript before, there are five other datasets in NCBI/ENA, of which we had not been aware previously. These are now also included in the manuscript.

In addition to the Reviewers' comments, I could add the following minor points:

- L41 Prasinodermophyta has been proposed as a third major lineage of Chloroplatida besides chlorophytes & streptophytes: https://www.nature.com/articles/s41559-020-1221-7
  >> The text has been edited accordingly.

- Fig. 2 I assume the two dots with lower GC% correspond to the mitochondrial and plastid genomes, as suggested in the caption. But could you indicate which is which?
  >> The figure and caption have been updated.

- L83 abbreviation for hour is h
  >> This has been corrected.

- L94 quantity and quality?
  >> This has been corrected.

*by **Iker Irisarri**, 22 May 2024 11:48*
Manuscript: **https://doi.org/10.1101/2023.07.11.548521**
version: 1

Review by anonymous reviewer 1, 30 Apr 2024 05:51

The authors presented a high-quality assembly genome of microalga *Coccomyxa viridis*, and did the annotation. This manuscript provides useful resources of microalgae. I have some questions on the manuscript.

- To evaluate the completeness of genome assembly, do the authors perform the genome size estimation of the microalgal based on experimental and computational method?
  >> In this genome note, we present the assembly of the *C. viridis* SAG 216-4 genome based on PacBio Hifi, ONT, and HiC sequencing. Together, the data resulting from using these technologies led to a chromosome-scale assembly consisting of 19 nuclear, one plastid and one mitochondrial scaffold. As described in the manuscript, the total length is 50.9 Mb. Here it needs to be noted, that genome sequencing (PacBio Hifi, ONT) and HiC sequencing are completely independent, orthogonal analyses that confirm each other's outcome. Following annotation, the completeness of the nuclear genome assembly was determined to be very high with a BUSCO score of 98.6% (chlorophyte_odb10 database).

- The authors showed that the assembly is chromosome-scale level, I wonder if the authors have any data on the chromosome number of this alga.
  >> We carried out HiC sequencing to manually curate our assembly and to determine the number chromosomes. Our analyses resulted in the identification of 19 scaffolds (Figure 1b). Most of these scaffolds carry telomeric repeats at both ends, suggesting they are complete chromosomes. Two additional scaffolds represent the mitochondrial and plastid genomes, respectively. This is in accordance with previously published data (https://link.springer.com/article/10.1186/gb-2012-13-5-r39).

- Line 202-205, the authors gave conclusion that scaffold 20 and 21 are chloroplast and mitochondrial genomes, these just only based on the length and GC content, I think it may be not correct, same as the conclusion in Figure 1a legend. Did the author map the scaffolds to reference plastome and mitogenome?
  >> In contrast to scaffolds 1-19, scaffolds 20 and 21 were considerably shorter and displayed a markedly lower GC content. While there are some green algae with a bias toward GC in their plastid and mitochondrial genomes (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0023624), this does not seem to be the case for *C. viridis* as shown in this manuscript and elsewhere (https://www.nature.com/articles/s41598-023-48637-w). To confirm that scaffolds 20 and 21 do represent the chloroplast and mitochondrial genomes, we additionally carried out BLAST analyses, which identified plastid and mitochondrial genes on the respective plastids. We have now also annotated those genomes as shown in the new Figure 2. The corresponding data have been uploaded to ENA and can be found under the project accession number PRJNA1054215.

- I want to ask if the authors have examined the scaffold 1-19 containing any plastome or mitogenome fragments?
  >> As mentioned above, the plastid and mitochondrial genome assemblies have now been annotated and are available on ENA. Due to the contiguity of the nuclear

chromosomes represented by scaffolds 1-19, we have not additionally determined whether they contain plastome or mitogenome fragments.

Title and abstract
Does the title clearly reflect the content of the article? **Yes**
Does the abstract present the main findings of the study? **Yes**
Introduction
Are the research questions/hypotheses/predictions clearly presented? **Yes**
Does the introduction build on relevant research in the field? **Yes**
Materials and methods
Are the methods and analyses sufficiently detailed to allow replication by other researchers? **Yes**
Are the methods and statistical analyses appropriate and well described? **Yes**
Results
In the case of negative results, is there a statistical power analysis (or an adequate Bayesian analysis or equivalence testing)? **Yes**
Are the results described and interpreted correctly? **Yes**
Discussion
Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument? **Yes**
Are the conclusions adequately supported by the results (without overstating the implications of the findings)? **Yes**

Review by Elisa Goldbecker, 02 May 2024 16:20

Kraege et al. provide the first genome of the chlorophyte and lichen photobiont *Coccomyxa viridis* (SAG 216-4). They generated a high-quality assembly using long-reads by PacBio-HiFi and Oxford Nanopore, that were scaffolded using Hi-C. The assembly was further annotated using RepeatMasker and Braker software. The paper outline is very clear and concise. I will not comment on assembly methods, as this falls outside of my expertise. However, I have some small remarks regarding general things and the annotation:

- Introduction: Terms such as “early diverging” (line 43) should be avoided as they can lead to false tree thinking. (McDaniel, 2021), https://doi.org/10.1111/nph.17241
  >> We thank the reviewer for their comment. The term “early diverging” has been removed from the manuscript.

- Methods: RNAseq; It is not mentioned how many RNAseq samples were generated.
  >> Total RNA was extracted from a single dense 9-day-old algal culture as described in the “DNA and RNA extraction” section. 500 ng of this total RNA was then used to generate the sequencing library as described in the paragraph on “RNA sequencing”.

- Annotation: It is stated that BRAKER was run using transcriptome evidence only, however BRAKER2 is cited, which describes the implementation of BRAKER using protein data. The citation should be changed to BRAKER1 e.g. Hoff et al. 2016 https://doi.org/10.1093/bioinformatics/btv661
  >> The reference has been changed accordingly.

- Results: The claim that the average level of alternative splicing is predicted to be very low is in my opinion too speculative, as apparently only RNAseq data from one condition was used and also the number of RNAseq samples is unknown.
  >> We agree with the reviewer that our analyses do not provide definitive proof of a low level of alternative splicing. In line 235, the text carefully states that this is a prediction based on the estimated number of transcripts per gene (see also Table 1). However, we have added a sentence stating that further analyses are required to confirm the actual amount of alternative splicing,

- Data availability: Data should be made available upon publishing.
  >> Data are available on ENA under the study accession number PRJNA1054215.

Review by Fabian Haas, 09 May 2024 14:34

Review Kraege et al.
In this manuscript, 'High quality genome assembly and annotation (v1) of the eukaryotic terrestrial microalga *Coccomyxa viridis* SAG 216-4' posted July 12, 2023 at bioRxiv, the authors present the first fully assembled genome of the eukaryotic terrestrial microalga *Coccomyxa viridis* SAG 216-4. Besides the genome assembly the authors performed repeat masking, gene annotation, contamination analysis, synteny detection, and a ploidy test.

The manuscript presents the resource of the genome and is kept technical. I'm missing the biological meaning and some more analyses. At the introduction the authors are asking the question of the molecular mechanisms that determine the various symbiotic lifestyles. The manuscript does not show the approach to answer this question. E.g. the article published by Tagirdzhanova et al., 2023 (Sci Rep), uses, among other things, the genome assembly by Kraege et al. and shows some more biological context. Is there any gene loss or gene transfer at *Coccomyxa viridis* compared to free living *Coccomyxa* species?

>> We thank the reviewer for this comment. This manuscript is indeed very technical and is inteneded as a genome note only. Further analyses addressing biological questions are beyond its scope but will certainly be addressed elsewhere. However, we have now included the annotations of the plastid and mitochondrial genomes (new Figure 2). Additionally, we have clarified the intention behind this manuscript in the introductory section.

Suggestions of additional analyses for this paper with the existing dataset:

- Hi-C: The telomere boundaries were mentioned. What about centromeres? Are there TADs or other structural elements or A/B compartments? Is the Hi-C resolution high enough to say anything about the 3D structure?
  >> The HiC map presented in Figure 1b indicates abundant contacts pointing toward a developed 3D structure. Additionally, for some chromosomes, centromeres are visible in the contact map. This is now indicated in the text. Any additional analyses proposed by the reviewer are interesting, yet beyond the scope of this genome note.

- Nanopore (ONT): The ONT data can be used to detect methylation (e.g. 6mA or 5mC). https://github.com/nanoporetech/dorado
  >> Also this analysis is beyond the scope of this manuscript.

- RNA-seq: Are there alternative splicing sides, start codons, rDNA arrays?
  >> Based on our analysis, we observe, on average, one transcript per gene model, suggesting a low amount of alternative splicing as indicated in lines 241-243 and in Table 2.

- Assembly: Does the assembly contain endogenous viral element(s)? Are there any interesting TE structures like the Chlorella zepp retro TE at the centromere? Are there sub-telomere structures or TEs at the telomeres?
  >> As mentioned earlier, for some chromosomes centromeres are visible in the HiC contact map. However, based on GC content and repeat location, we cannot

identify all centromeric regions. A full characterization would require ChiP-seq for CenH3 mapping as has been done for the red alga *Cyanidioschyzon merolae* (e.g. https://www.sciencedirect.com/science/article/pii/S0014579315002471).

A few minor points:

Line 28: nineteen => 19 >> The text has been changed accordingly.

Line 81: 3x vitamins => which? >> The information has been added to the text.

Line 93/94: DNA quality and quality => quantity >> The text has been changed accordingly.

Line 109: Why was the Rapid Sequencing Kit used? >> Our lab frequently sequences whole genomes from various organisms using the ONT technology. So far, we have always obtained good results using the Rapid Sequencing Kit and have used it here as well, instead of the Ligation Kit.

Line 111: Flow Cell 9.4.1 => which device? >> We have used a minION device. This information has now been added to the text.

Line 146: Why manually at the first place? Who many gaps were left after Hi-C? Usually, ARCS (doi:10.1093/bioinformatics/btx67) or TGS-gapcloser (doi:10.1093/gigascience/giaa094) are performing well. >> Genome assembly with the PacBio Hifi reads using Raven resulted in 27 contigs. During the manual curation and scaffolding based on the HiC data, some of these contigs were rearranged resulting in 15 gaps between contigs on 21 scaffolds. Using the ONT data, we could easily close most (9/15) of these gaps manually.

Line 175: Were protein files of other green algae included at the braker run or only the RNA-seq bam files? >> Gene annotation has been carried out based on the RNA sequencing data only. This is now more clearly stated in the manuscript.

Title and abstract
Does the title clearly reflect the content of the article? **Yes**
Does the abstract present the main findings of the study? **Yes**
Introduction
Are the research questions/hypotheses/predictions clearly presented? **No** – The history and differentiation of *Coccomyxa* was shown. And the question of the molecular mechanisms that determine the various symbiotic lifestyles was asked. I'm missing a clear statement how this new genome assembly will help answering this question.

>> As stated above, the introduction has been edited accordingly and now clearly states that this manuscript presents the genome assembly and annotation of *C. viridis* SAG 216-4, which will facilitate the investigation into symbiont-related traits and their evolution among *Coccomyxa* spp. in the future.

Does the introduction build on relevant research in the field? **Yes**
Materials and methods

Are the methods and analyses sufficiently detailed to allow replication by other researchers? **Yes**
Are the methods and statistical analyses appropriate and well described? **Yes**

Results

In the case of negative results, is there a statistical power analysis (or an adequate Bayesian analysis or equivalence testing)? No negative results
Are the results described and interpreted correctly? **Yes**

Discussion

Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument? **No** – The results are representing not everything the data could show. Some analyses are missing.

> >> As stated above and in the text, this manuscript represents a genome note and any additional comparative genomics analyses are beyond its scope.

Are the conclusions adequately supported by the results (without overstating the implications of the findings)? **Yes**

Review by anonymous reviewer 2, 19 May 2024 05:53

The aim of this manuscript was to create a high-quality reference genome for the *Coccomyxa viridis* SAG 216-4 strain. The genome is a valuable resource for future studies on this ecologically widespread and versatile fungal lineage.

The analyses performed by the authors are of a high standard. However, I am missing a context as to why this data should be published as a scientific article rather than as a genome report. Except for the synteny graphs, all the data shown by the authors will be available upon their release on ENA.

>> We thank the reviewer for their evaluation. We have modified the introductory section to clarify that this manuscript is a genome report only and, therefore, does not address any biological questions. Given its high-quality and contiguity, we regard the generated assembly a valuable resource to address the origin and evolution of symbiotic traits in *Coccomyxa* spp. in the future.

The authors mentioned at lines 55-57 that "*Coccomyxa* and a genome sequence is available only for a single *Coccomyxa* species, namely *Coccomyxa subellipsoidea* C-169, which was isolated in Antarctica where it occurred on dried algal peat (Blanc et al., 2012)." However, I found seven *Coccomyxa* genomes on NCBI and two more on ENA. I consider that all these genomes should be taken into account and, in my opinion, even included in the analyses. This is because the authors already compared their strain with *Coccomyxa subellipsoidea* C-169. Given that more genomes are available, the authors should either include these additional genomes in their analyses or provide a new rationale for focusing solely on the comparison with *C. subellipsoidea* C-169.

>> The only other high-quality genome of *C. viridis* has been made available following the upload of the first version of this manuscript to BioRxiv and is now referred to in the text. In addition to the genome of *C. subellipsoidea* C-169, which was included in this manuscript before, there are five other datasets in NCBI/ENA, of which we had not been aware previously. These are now also included in the manuscript.

Furthermore, since the comparison between strains was already initiated, I recommend conducting additional comparative analyses to match those performed for their strain, such as GC content, genome size, genome completeness, and the number of genes (as shown in Table 2). Does *C. subellipsoidea* C-169 also have signal proteins? Is the genome contaminated, etc. If the authors decide to consider all the strains from ENA and NCBI, I suggest they also create a phylogeny using the Busco genes.

>> We agree with the reviewer that the suggested analyses would result in interesting and valuable insights into *Coccomyxa* biology. However, since this is not a scientific article but a genome report, we regard these analyes to be beyond the scope of this manuscript.

Furthermore, I would like the authors to address a few minor details:

- Early diverging" is an incorrect term to refer to sister clades. While it is commonly used, the term inaccurately implies that these lineages evolved earlier than their sister groups. In reality, all extant lineages have evolved over the same amount

of time, and no lineage is older than another. For a more detailed explanation, I recommend reading this blog post: [The Ancestors Are Not Among Us.]
>> We thank the reviewer for their comment. The term “early diverging” has been removed from the manuscript.

- Lines 182-184: ‘All software and tools used for the genome assembly and annotation are summarized in Table 1’. This should be a supplementary table and the genome stats should be the Table 1.
  >> The original Table 1 is now Supplementary Table 1. The other Tables have been renumbered and are referred to in the text accordingly.

- 239-240 ‘BLAST analyses of six identified ITS sequences in the *C. viridis* SAG 216-4 assembly confirmed its species identity.’ I am curious if all ITS copies were identical. This is interesting to know and be documented.
  >> The identified ITS sequences were not identical as shown in the alignment below. Still, the first 5 ITS sequences identify this strain SAG 216-4 as *C. viridis*, and the last ITS sequence also confirmed its identity as *Coccomyxa*. The text has been edited accordingly.

```
HiC_scaffold_2_-_1149695_1155536|F|ITS1      CATCGATCAAACACCTTTCCACGCGAACACTCGTCTCCCCGCGCTGCTGTTTTGATTGGG      60
HiC_scaffold_2_+_1066943_1072782|F|ITS1      CATCGATCAAACACCTTTCCACGCGAACACTCGTCTCCCCGCGCTGCTGTTTTGATTGGG      60
HiC_scaffold_2_-_1468241_1474069|F|ITS1      CATCGATCAAACACCTTTCCACGCGAACACTCGTCTCCCCGCGCTGT-CTTTC----TGA      55
HiC_scaffold_2_-_1274203_1285030|F|ITS1      CATCGATCAAACACCTTTCCACGCGAACACTCGTCTCCCCGCGCTGT-CTTTC----TGA      55
HiC_scaffold_1_-_2416277_2422105|F|ITS1      CATCGATCAAACACCTTTCCACGCGAACACTCGTCTCCCCGCGCTGT-CTTTC----TGA      55
HiC_scaffold_1_+_2466038_2471866|F|ITS1      CATCGATCAAACACCTTTCCACGCGAACACTCGTCTCCCCGCGCTGT-CTTTC----TGA      55
HiC_scaffold_2_+_947009_952839|F|ITS1        CATCGATCAAACACCTTTCCACGCGAACACCCGTCTCCCCGCGCTGTTCTTTC----TGA      56
                                             ****************************** **************   ***     *.

HiC_scaffold_2_-_1149695_1155536|F|ITS1      ATAGCTTATGACGTCCTCTCTGGAGGAAAGAGGCTTTTCCCATCTGGTAGAAACGTGCGC      120
HiC_scaffold_2_+_1066943_1072782|F|ITS1      ATAGCTTATGACGTCCTCTCTGGAGGAAAGAGGCTTTTCCCATCTGGTAGAAACGTGCGC      120
HiC_scaffold_2_-_1468241_1474069|F|ITS1      TCGGCTTATGACGTCCTCTCCCGAGGAAAGAGGCGAT-------CGAAAGGAGCGTGCGC      108
HiC_scaffold_2_-_1274203_1285030|F|ITS1      TCGGCTTATGACGTCCTCTCCCGAGGAAAGAGGCGAT-------CGAAAGGAGCGTGCGC      108
HiC_scaffold_1_-_2416277_2422105|F|ITS1      TCGGCTTATGACGTCCTCTCCCGAGGAAAGAGGCGAT-------CGAAAGGAGCGTGCGC      108
HiC_scaffold_1_+_2466038_2471866|F|ITS1      TCGGCTTATGACGTCCTCTCCCGAGGAAAGAGGCGAT-------CGAAAGGAGCGTGCGC      108
HiC_scaffold_2_+_947009_952839|F|ITS1        TCGGCTCATGACGTCCTCTCCGGAGGAAAGAGGTGAT-------CGTAAGACGCGTGCTC      109
                                             : .*** *************  ***********  :*        * :**...***** *

HiC_scaffold_2_-_1149695_1155536|F|ITS1      CTGCTCCCGGTCGACTAGCTCCGGCCAGTCGGCAAGGTCAGGCGGTGCCGTCCGAGGGAT      180
HiC_scaffold_2_+_1066943_1072782|F|ITS1      CTGCTCCCGGTCGACTAGCTCCGGCCAGTCGGCAAGGTCAGGCGGTGCCGTCCGAGGGAT      180
HiC_scaffold_2_-_1468241_1474069|F|ITS1      CTGCTCCCGGTCGACTGGCTCAGGTCAGTCGGCAAGGTCAGGCGGTGCCGTCCGAGAGAT      168
HiC_scaffold_2_-_1274203_1285030|F|ITS1      CTGCTCCCGGTCGATTGGCTCAGGTCAGTCGGCAAGGTCAGGCGGTGCCGTCCGAGGGAT      168
HiC_scaffold_1_-_2416277_2422105|F|ITS1      CTGCTCCCGGTCGACTGGCTCAGGTCAGTCGGCAAGGTCAGGCGGTGCCGTCCGAGAGAT      168
HiC_scaffold_1_+_2466038_2471866|F|ITS1      CTGCTCCCGGTCGACTGGCTCAGGTCAGTCGGCAAGGTCAGGCGGTGCCGTCCGAGAGAT      168
HiC_scaffold_2_+_947009_952839|F|ITS1        CTGCTCCCGGTCGATTGGCTCCGGTCAGTCGGCAAGGTCAGGCGGTGCCGTCCGAGGGAT      169
                                             ************** *.****.** ********************************.***

HiC_scaffold_2_-_1149695_1155536|F|ITS1      GAGGCTCTCTCGAGAGCCGACGACTCCCGCGGCTCCGGCTGGGACCAAACTCCAACCGAT      240
HiC_scaffold_2_+_1066943_1072782|F|ITS1      GAGGCTCTCTCGAGAGCCGACGACTCCCGCGGCTCCGGCTGGGACCAAACTCCAACCGAT      240
HiC_scaffold_2_-_1468241_1474069|F|ITS1      GAGGCTCTCTCGAGAGCCGACGACTCCCGCGGCTCCGGCTGGGTCCAAACCCCAACCGAT      228
HiC_scaffold_2_-_1274203_1285030|F|ITS1      GAGGCTCTCTCGAGAGCCGACGACTCCCGCGGCTCCGGCTGGGACCAAACCCCAACCGAT      228
HiC_scaffold_1_-_2416277_2422105|F|ITS1      GAGGCTCTCTCGAGAGCCGACGACTCCCGCGGCTCCGGCTGGGTCCAAATCCCAACCGAT      228
HiC_scaffold_1_+_2466038_2471866|F|ITS1      GAGGCTCTCTCGAGAGCCGACGACTCCCGCGGCTCCGGCTGGGTCCAAACCCCAACCGAT      228
HiC_scaffold_2_+_947009_952839|F|ITS1        GAGGCTCCCGCGAGAGCCGACGACTCCCGCGGCTCCGGCTGGGACCAAACCCCAACCGAT      229
                                             ******* * ***********************************:*****  *********

HiC_scaffold_2_-_1149695_1155536|F|ITS1      CCTTGACGCATCAAATCTGAAGCCGAGCGCACGCCGACCTCGAAAGAGACGGCGCTGCCT      300
HiC_scaffold_2_+_1066943_1072782|F|ITS1      CCTTGACGCATCAAATCTGAAGCCGAGCGCACGCCGACCTCGAAAGAGGCGGCGCTGCCT      300
HiC_scaffold_2_-_1468241_1474069|F|ITS1      CCTTGACGCATCAAATCTGAAGCCGAGCGCACGCCGGCCTCGAAAGAGACGGCGCTGCCT      288
HiC_scaffold_2_-_1274203_1285030|F|ITS1      CCTTGACGCATCAAATCTGAAGCCGAGCGCACGCCGGCCTCGAAAGAGACGGCGCTGCCT      288
HiC_scaffold_1_-_2416277_2422105|F|ITS1      CCTTGACGCATCAAATCTGAAGCCGAGCGCACGCCGGCCTCGAAAGAGACGGCGCTGCCT      288
HiC_scaffold_1_+_2466038_2471866|F|ITS1      CCTTGACGCATCAAATCTGAAGCCGAGCGCACGCCGGCCTCGAAAGAGACGGCGCTGCCT      288
HiC_scaffold_2_+_947009_952839|F|ITS1        CCTTGACGCATCAAATCTGAAGCCGAGCGCAGGCCGGCCTCGAAAGAGACGGCGCTGCCT      289
                                             ******************************** ****.***********.***********

HiC_scaffold_2_-_1149695_1155536|F|ITS1      TCAAACCAAAGA       312
HiC_scaffold_2_+_1066943_1072782|F|ITS1      TCAAACCAAAGA       312
HiC_scaffold_2_-_1468241_1474069|F|ITS1      TCAAACCAAAGA       300
HiC_scaffold_2_-_1274203_1285030|F|ITS1      TCAAACCAAAGA       300
HiC_scaffold_1_-_2416277_2422105|F|ITS1      TCAAACCAAAGA       300
HiC_scaffold_1_+_2466038_2471866|F|ITS1      TCAAACCAAAGA       300
HiC_scaffold_2_+_947009_952839|F|ITS1        TCAAACCAAAGA       301
                                             ************
```

- About the KAT plots: Based on the KAT graph shape I see that the genome is haploid. Maybe is good to say this in the manuscript but also do KAT plots on the other genome/genomes.
  >> Yes, the genome of *C. viridis* SAG 216-4 is haploid as stated in the caption of the new Figure 4 and in line 234.