INRAE Grignon, 12th of September 2022

Dear Recommender and Reviewers,

Thanks to the comments of the reviewers, we are happy to provide a revised version of our manuscript. We believed we answered to all the questions and suggestions. Our answers are inserted below the comments of the two reviewers (*in italics*).

Best regards,

Adeline Simon and Muriel Viaud

Reviews

Reviewed by Cecile Lorrain, 13 Apr 2022 11:10

I have two main recommendations to improve the manuscript:
1- To further support the presence of accessory chromosomes in the newly sequenced strains, the authors could perform a pulsed-field gel electrophoresis. This potentially would compensate for the scarcity of telomeric repeats in some chromosomes.

Response:

Thank you for this relevant suggestion. We managed to access to the required pulsed-field gel electrophoresis equipment, to isolate intact chromosomal DNA and to separate the smaller chromosomes (< 1 Mb) of the B05.10, Sl3 and Vv3 strains (Sup. Fig. S2). These new data confirmed the presence of the ACs expected from the assembly data of the genomic sequences (mentioned Page 6-7, Lines 123-128). The methodology was added in the Material and Methods section (Page 27, lines 689-693).

2- RIP and GC content: even though RIPped sequenced have low GC content, GC content alone is not the best signature for RIP, since many species exhibit TEs with low %GC but do not have RIP. It can reflect the tendency of TEs to accumulate mutations over time. I suggest the authors, if they want to emphasize the presence of RIP signatures on TEs, to run The RIPper (https://theripper.hawk.rocks/#/home). This user-friendly tool search for RIP indexes in whole-genome or TEs sequences. Also, authors could search for the presence of DNMTs (especially dim2 and RID) genes which are know to be involved in RIP in fungi.

Response:

We agree that a low % of GC alone is not sufficient to prove the occurrence of RIP. Actually, what we only wanted to say in the manuscript is that a high % of GC indicates the absence of RIP. As all the complete copies of the newly identified TEs that produce small RNAs (Copia_4,

Gypsy_6 and Gypsy_7 in Vv3) have high GC contents, we concluded they are not affected by RIP.

*The presence of RIP signatures in the genome of B. cinerea was fully investigated in the study of Amselem et al. (2015, BMC Genomics **16**: 141) and we should have mentioned it better in our initial manuscript. In this previous work, RIP signatures (both at CpA and CpT dinucleotides) were found in TEs of B. cinerea and were indeed associated with the presence of two genes encoding the cytosine DNA methyltransferases, known to be required for RIP (BcRID1 (Bcin03g04600) and BcRID2 (Bcin09g05050). More recently, other studies have shown signatures of RIP in TEs of different strains of B. cinerea using the RIPcal tool (Porquier et al., 2016; 2019; 2022).*

As suggested, we tested the RIPper tool. The results indicated comparable proportions of RIP-affected sequences in the B05.10, Sl3 and Vv3 genomes. Nevertheless, this tool only takes into account RIP occurring at CpA loci (i.e. the dominant form of RIP in Ascomycetes) but not RIP occurring at CpT loci that is also very important in B. cinerea (Amselem et al., 2015).

To summarize, in the revised version of the manuscript, we added a short description of the work of Amselem et al. (2015) and of Porquier et al. (2016; 2019; 2022) (Page 13, lines 293-298) and we tried to be more careful about the meaning of the % of GC (Page 13, lines 301-302).

Minor:
- In line 31-33, the authors state that generalists could actually be represented by different co-existing populations specialised in infecting different host species. It would be nice to have the demonstration that it is the case for *B. cinerea*, just right after the scenario statement. The demonstration comes only from lines 49-61. I would move this part line 36 and finish the paragraph with the different gene catalogues.

Response:

OK, we reorganized this part of the introduction as suggested.

-L421-423: remove sentence "Both inactive and active TEs can additionally generate genomic rearrangements by homologous recombination as discussed below."

Response:

OK, we removed this (unnecessary) sentence.

Reviewed by Thorsten Langner, 07 Apr 2022 13:42

I only have a few comments that I suggest should be addressed:

Page 15, line 304:

Here, the authors state that predominantly TEs with high GC content produce small RNAs and this appears again in the discussion, but the data is not shown. In my opinion this is an interesting finding which should be supported by data. Otherwise, it is hard to tell how clear the correlation between GC content and small RNA expression is, especially given that the majority of TEs appears to be in the >40%GC category (Figure S11). In addition, the spatial distribution of siRNA-producing TEs in the chromosome level assemblies could add additional value. I suggest to either show this type of analysis or to remove statements and conclusions that are not supported by the results.

Response:

The reason why we initially didn't show the correlation between the GC content of TEs and the production of small RNA expression is that this correlation was previously shown in the work of Porquier et al. (2021). However, we agreed that it would be useful to display our own results, and we therefore added a new figure that shows the mapping of the small RNAs reads on the complete copies of TEs in relation to their GC content (Sup. Fig. S15).

Regarding the spatial distribution of siRNA-producing TEs, Suppl Figure S17 actually displays the position of the three TE of interest i.e. Copia_4, Gypsy_6 and Gypsy_7 on the chromosomes of Vv3.

I also wonder if the age of TEs and extent of RIP influences the mapping quality of TE derived small RNAs to the consensus and the filtering/masking. This could potentially have an important impact on the overall conclusions of the small RNA producing TEs. Could the authors please comment which quality controls they have done? In the methods section it says, "Quality controls have been done, including PCA…", however it is hard to tell if any of these controls address the impact of mapping/filtering (e.g. how many reads that map to TEs were discarded during the filtering process? Among those reads, is there a bias that could influence conclusions about GC richness/small RNA expression?).

Response:

This is indeed an important point that needed to be clarified.

The quality controls and the processing of the small RNA reads is described in the Methods section (Page 29, lines 732-743. As indicated, the filtering steps are the same than in Weiberg et al. (2013). For more transparency, we added the number of reads remaining after each filtering step in the revised Sup. Table S4. We also better explained why we focused on 20-24nt reads (Page 14, lines 335-337 in results).

Furthermore, we performed, as an additional control, a mapping of the unprocessed (raw) reads of small RNAs on TE consensus sequences and complete copies of TEs. This mapping did not reveal any additional small RNA-producing TE and the results were therefore fully consistent with those obtained with the previously selected reads. This quality control was added in Material section (Page 29, lines 743-745).

Page 17, line 348 onwards:

The PCR results and the mapping results are somewhat inconsistent for some isolates. Mapping shows presence of gypsy6 across all tested G1 isolates (Vv1,Vv3,Vv5,Vv14) but PCR fails to detect gypsy6 in these strains. I think the PCR results should thus be interpreted with caution, as presence/absence analysis by PCR can sometimes be problematic. This discrepancy should at least be discussed. Additionally, there are only few details regarding the PCR assay in the paper. It would help if the authors could provide the template sequences used to design the primers and variation of TEs (e.g. alignment of identified TEs similar to what they have done in Figure S8 for the Boty TE, with focus on the primer binding sites). Could variation of TE sequences contribute to absence of PCR products? This could explain the differences between mapping and PCR detection.

Response:

Yes, we agree with the fact that the PCR assays should have been better described and that negative results should be interpreted with caution. As suggested, we added a new figure to show the alignments of the copies of TEs and the localization of the PCR primers (revised Sup. Fig. S18). In the revised manuscript, we also described the PCR results and we explored more carefully what could explain the discrepancy between the small RNAs mapping results and the PCR detection of Gypsy_6 in Vv1, Vv5 and Vv14 strains (Page 17, lines 422-437). As shown, in the part C of the revised Sup. Fig. S18, the lack of PCR product in Vv5 could be due to the absence of the regions where the primers have been designed. This hypothesis could not be raised for the Vv1 and Vv14 strains, and here we hypothesized that the lack of PCR product could be due to the absence of a Gypsy_6 copy containing both primer regions.

Minor comments:

Figure S3: Colors of lines and dots should be explained in the figure legend. Similarly, solid and dashed lines should be explained.

Response:

A legend was added (now Sup. Fig. S4).

Page 11, line 202: The abbreviation “IPR domains” should be spelled out.

Response:

OK, this was done (“InterProScan domains”).

Page 12, line 213: “Fifteen consensus were previously identified” should probably read ”…consensus sequences…”?

Response:

Yes, we added the word “sequences”.

Page 19, lines 388/390: These two sentences read a bit contradictory. Maybe it would be worth rephrasing this to point out the lineage specificity of the acc. Chr BCIN19 as a clear difference to 17/18 for which lineage specificity was not supported.

Response:

For more clarity, we modified these sentences (Page 19, lines 483-486).

Page 15, lines 291 and following"…producing the higher amount of small RNAs…" should be changed to "…highest amount.."

Response:

OK, this was changed.