Response to reviewers

Review by Grégoire Aubert

Q1: One point to clarify is the plant material that has been used to measure the severity of the symptoms, as it impacts the independence of this trait from incidence.

Plant material used to measure the severity of RHBV symptoms was different from the plants used to evaluate incidence. Severity and Incidence were evaluated in different experiments. In order to evaluate severity, 10 plants per F_3 family were measured individually. 18 days after sowing plants were in individual transparent tubes with 3 nymphs per plant. Nymphs were allowed to feed for three days on the plants. A separate experiment was conducted to evaluate incidence using 60 plants per F_3 family (3 replicates, 20 plants each replicate) in a randomized complete block design.

We completed the text of the paragraph *Severity* in Materials & Methods.

Q2: Title: I would suggest: Genetic bases of resistance to the rice hoja blanca disease deciphered by a QTL approach

We accept this suggestion, which gives more precision to the Title.

Q3: Line 62: Has the local ancestry analysis mentioned here been performed by the authors or published elsewhere? Please provide a reference.

Local ancestry analysis was performed by us, and the results are shown in Figure S2 (see line 375).

We added a description of the methodology in the legend of Figure S2.

Q4: Line 123: It is not totally clear to me on which plants the ALA is calculated: is it only on plants with symptoms identified in the previous step? Or is it on 10 random infected plants per F3 family? The mention of "plants with symptoms" is ambiguous, please clarify this point as it is important for the interpretation of the results. If done on 10 random plants, your severity measure depends strongly of the incidence rate. If done only on plants with symptoms, it would help to disconnect incidence from severity of the symptoms, and dissect the mechanisms of resistance.

ALA was estimated in 10 plants one by one in an independent experiment as described in the answer to question 1. Those plants that did not show symptoms were scored as a 0% of ALA and were included to calculate the average of ALA per F_3 family.

We added an explanation in the paragraph Severity in Materials & Methods.

Q5: Line 142: A supplemental table with the markers used in the study would be helpful (SNP flanking sequence, genome position, parental alleles) to allow replication by other researchers. That would also be particularly useful for breeders for marker-assisted selection.

We prepared the table (Table S1) with SNP flanking sequence, genome position, and parental alleles and added it as supplementary data.

We added a phrase in Results / SNP markers and genetic maps to refer to the table.

Q6: Line 204: 'following' instead of 'flowing'?

Yes, thank you.

Q7: Line 232: A supplemental figure illustrating the segregation distortion along the genome for the 4 crosses would support the text well.

We added a supplemental Figure S1 that shows departure from expected genotypic frequencies in the chromosomes listed in the paragraph. Previous Figure S1 is thus now Figure S2, and previous Figure S2 is thus now Figure S3.

Q8: Line 251:" a significant lower severity": Considering the confidence interval (25.8+-32.7%) given for WAS208 for severity, are the F3 families with low severity marks really significantly different than WAS208? What threshold has been chosen? How many families are in this case?

We agree with the reviewer and removed the sentence where we state that some F_3 families exhibited a significant lower severity than the resistant parent.

Q9: Line 271: If severity has been measured on all plants, and not only on symptomatic plants, the interdependence between the two traits is clear.

Indeed, estimation of severity included all plants, even those with no symptoms. This thus creates correlation between the two traits.

We re-calculated all the QTLs using ALA_strict, which does not include the plants without symptoms in the averaged ALA.

Q10: Line 309: I am not sure about the conclusion about qHBV11.1 controlling "primarily" the RHB severity in FD2000, infection being prior to the expansion of symptoms. This QTL also explains a substantial part of incidence in that cross (R²=0.29).
It would be interesting to find out why no effect of the QTL on incidence was found in WAS208 and Badka (Detection threshold? Interactions with other QTLs? Level of infection lower in the FD2000 cross?)

We agree that our conclusions regarding the control of severity vs incidence might be somehow confusing. We initially concluded that qHBV11.1 controls primarily RHB severity because the LOD score is higher for severity than for incidence in Fd2000: since we included the plants with no symptoms in the QTL detection analyses, incidence was correlated with severity by mathematical construction, inducing a QTL just by correlation. In Was208 and Badka, there is no QTL for incidence in the region, so the QTL is clearly a severity QTL. However, in Fd2000, it might be that a QTL for incidence is also present in Fd2000.

We changed the interpretation in the text.

Q11: Line 393: On the potential role of AGO4 in virus resistance, I suggest to cite also Bhattacharjee et al. ("Virus resistance induced by NB–LRR proteins involves Argonaute4-dependent translational control." The Plant Journal 58.6 (2009): 940-951).

This is a great suggestion.

We added a phrase that cites Bhattacharjee et al. Thanks very much.

Q12: Line 462: I didn't find any mention of the QTLs qHBV9 and qHBV10 in the tables or in the results section, they appear only in the discussion, am I right? Can it be added?

qHBV9 and qHBV10 are two minor QTLs located on chromosomes 9 and 10, respectively. The were detected only with CIM method. Their resistant allele comes from the susceptible parent

(Bluebonnet 50), so we assumed they might be susceptibility QTLs. Since we don't have enough evidence about those QTLs, we decided to exclude them from the analysis of this paper and to try to understand the role of these genomic regions more deeply in future experiments.

We thus removed the phrase that refers to them in the discussion.

Q13: Line 493: One application of this study is the possibility for breeders to increase the level of RHBV resistance by pyramiding the different QTL. That can be mentioned here.
Thanks very much for the suggestion.

We added a phrase in this sense to the Conclusion.

Review by V. Geoffroy

Q1: is not clear which genome sequence was used for the candidate gene identification; is this genotype resistant or susceptible. In other words, do you have a resistant or a susceptible allele?

The RefSeq used was that of Nipponbare v.7 (IRGSP v1), a temperate Japonica accession. to align the reads from the parental lines, variant detection, and annotation. Therefore, the physical position of the SNPs was determined relative to this reference genome. This accession is of intermediate phenotype. The candidate genes were identified based on their predicted or verified function, from which we could not push further to susceptible or resistant allele functions. So when we mentioned "resistant" or "susceptible" alleles, we are referring to the allele from the resistant or susceptible parent, respectively.

Q2: the authors should mention that the sequence of the target region in the resistant genotypes should be obtained to identify the resistant allele(s) (or gene). Indeed, Copy Number Variation (CNV) have often been observed at resistance loci, and the resistance gene might not be present in the reference genome.

We agree that CNVs could be explaining the resistant phenotype. It is also possible that the responsible gene for the resistance is not present in the reference genome, however, good candidate genes for both qHBV4.1 and qHBV11.1 are present in the reference genome. We suggested that there are genetic variants (SNP, small Indels, or a CNV) either inside the resistance gene or close to it associated with the phenotypic variation. In a future study, we will characterize the candidate genes located in this region and analyze the structural variation that can be present in the resistant genotypes.

We added a phrase in the Conclusion (line 502) to follow the suggestion.

Q3: L371-373: Can you clarify this sentence: "In addition, the little variation in nucleotides — less than 1% — in the region of qHBV4.1, between Badka and PTB 25 indicates a probable common local ancestry (Figure S2). " Where did you get the sequence data of these two genotypes? What do you mean by "a probable common local ancestry". I would suggest to reformulate this idea, thanks.

This is also Q3 of Reviewer 1. Local ancestry analysis was performed by us, and the results are shown in Figure S2 (see line 375).

We added a brief description of the methodology in the legend of Figure S2.

Q4: Figure S2: sorry, this Figure is not clear. Could you clarify it, thanks. What is in abscissa and ordinate. Is there a reference associated. What do you mean by "probable local ancestry"?

We improved the legend and hope this makes the interpretation of the figure easier.

Q5: L 319: can you explain/develop what is the "joint and meta-analyses approaches" and/or add a reference?

Yes, we added an explanation in the *QTL mapping* section of Materials & Methods (L 212-231).

*Q7: Introduction L62-69: this paper is clearly not restricted to an audience working on rice, so could the authors develop/explain/put reference about indica/japonica;
Similarly, L62, "local ancestry analysis" is not obvious for non-expert audience.*

We introduced the *indica / japonica* classification, and briefly explained the principle of local ancestry analysis, with a reference (Santos et al 2019).

Q8: Miscellaneous
-Abstract: L21, no "s" at "symptoms severity" and L29, no "s" at "candidate genes identification". In general, the MS could benefit from English revision.
→ corrected

-L230: what do you mean by "level of purity"?
indicating the good level of purity of the F2 populations
→ changed to "indicating the absence of pollen contamination in the self-pollination process".

-L362: I suggest to replace "immunity" with "complete resistance"
→ corrected

-L384-396 : please indicate the % of the phenotypic variation explained by the QTL.
→ done (L403)

-L389:"putative gene" should be "putative candidate gene.
→ corrected

-L430: I suggest to use "gene" instead of "loci"
→ corrected

Table 1: precise the type of segregating population
→ done

Table 3, 4 and 5: problem in the presentation (for example "Dominance" written on 2 lines).
→ corrected