The line numbers indicated in our response to the reviewers' comments correspond to the corrected version with track changes.

Reviewer 1

In this manuscript, the authors argue that the evolution of virulence in the French wheat leaf rust population was driven by changes in the aggressiveness of the pathogen. They employed relevant experiments to achieve the goals of the study, used appropriate experimental designs and statistical procedures to analyze the data and interpret the results, and wrote the paper well.

In the introduction section and throughout the manuscript, the authors clearly defined the difference between virulence and aggressiveness. However, I would tend to disagree that virulence is purely qualitative and that the function of Avr proteins is solely for recognition by the plant R proteins. Several recent studies on rusts and other pathogens revealed that Avr proteins or 'effectors' have diverse functions, including blocking and manipulating host defenses. Any of them can be recognized by plant R proteins, which would trigger a downstream signaling response leading to effector-triggered immunity. In this sense, I view virulence in a more quantitative perspective and believe that the aggressiveness phenotypes described in the study are just manifestations of the other functions of effectors. In flax rust and oat crown rust, hundreds of predicted effectors were identified and they could have different functions.

→ We completely agree with Reviewer 1 that the same effector genes can have opposite functions: "promote infection" or "lead to effector-triggered immunity". This is precisely why we need distinct vocabulary to distinguish between both functions, and why we defined virulence as "the capacity of the pathogen to infect its host, in opposition to avirulence (qualitative)", and aggressiveness as "the quantitative variation of pathogenicity on a compatible host (quantitative)". These definitions match with the current, consensual use of vocabulary in plant pathology, even the debate about this issue is recurrent since some decades (e.g. Shaner et al. 1992 doi.org/10.1146/annurev.py.30.090192.000403 versus Andrivon 1993 doi:10.1094/Phyto-83-889). Virulence and aggressiveness refer to different phenotypic outcomes of the interaction without questioning the dual function of effector genes. Using the same word "virulence" for both functions would easily lead to misunderstanding, so we would like to keep using virulence and aggressiveness following the above definitions.

In the methods section, the authors conducted intricate procedures to measure pathogen aggressiveness. I was wondering about the method of extracting DNA from single pustules. Was DNA extracted from one single pustule on a leaf or were the spores from single pustule increased to sufficient amounts before extraction? If it's the former, I wonder about the non-specific binding of the primers to plant DNA during genotyping.

→ DNA was extracted from one single pustule on a leaf (containing no other pustule). The primers have been used in several other studies (Goyeau et al., 2007, 2012; Kolmer et al.

2019) and have proven to be specifically binding to Puccinia triticina, and not to wheat DNA. Moreover, in our study, we did not observe multiple bands after PCR amplification, which would be expected in case of multiple bindings.

In the results sections, the authors presented the increase in frequencies of new genotypes within a pathotype. Again, I would view this as a result of evolution of effectors within a race in response to the resistance genes present in the field. This usually happens in rusts as they can have multiple ways to generate genetic variability.

→ *In our study, new genotypes are only defined after genotyping with the 19 microsatellite markers. It does not allow us to speculate on the evolution of effector genes within a pathotype.*

In the discussion, I think it would be better to mention less of the results. The authors also mentioned that quantitative resistance leads to more aggressive isolates, which is in contrast with what we observe with other rusts. Could the authors expound more about the potential mechanism behind this?

→ *We agree with Reviewer 1 that quantitative resistance of the host and aggressiveness of the pathogen are two phenomena that are not necessarily linked. Consequently, we removed this part of the Discussion. We also shortened the last subsection of the Discussion.*

I believe qualitative resistance imposes a stronger selection pressure because it induces hypersensitive response or 'complete resistance' which is a very effective immunity reaction, thus requiring the pathogen to evolve in order to survive. Also, I disagree with the new term 'pathogenotype' because it would create a confusion with the existing meanings of pathotype and genotype. I think 'lineage' would be enough to describe genotypes within a pathotype.

→ *We appreciate the discussion about this point. We found the term 'pathogenotype' useful to rapidly distinguish between groups of isolates having a unique association between one pathotype and one genotype. The word 'lineage' suggests that all isolates having the same pathotype are related, which is not correct. Isolates with the same pathotype may have different origins. Thus, as this term is not wrong per se, we propose to keep it at the end of our discussion and to let the scientific community decide of its usefulness in the future. Nevertheless, we have improved the discussion by mentioning that "differentiation (or not) for neutral markers is independent from differentiation (or not) for functional mutations", line 671-672.*

I like the outlook of using genomewide markers to characterize leaf rust populations in the future. With the availability of a *P. triticina* reference genome, several isolates can be genotyped and association studies can be conducted to map genes responsible for the aggressiveness phenotypes, especially that the authors have a very good system in place for phenotyping.

→ *Thank you for your comment and appreciation.*

Reviewer 2

This is a very interesting manuscript about genotype dynamics and characterisation within two pathotypes of the wheat rust *P. triticina*. The authors have mined a rich, decades-long survey collection of *P. triticina* isolates that allows them to explore dynamics of pathotypes and genotypes within pathotypes over time. They have identified two major pathotypes for exploration of genetic and phenotypic variation. They identified a pair of genotypes that decreased/increased in frequency recently, for both pathotypes, and measured three components of aggressiveness for the losing and the winning genotypes to test the hypothesis that the winning genotype would show higher aggressiveness pathotype (i.e., more efficient at infecting, more rapid at sporulating and producing more abundant spores) on common wheat varieties they are able to infect.

→ We thank Reviewer 2 for his positive comments and for his interest in our work.

This may be the case for one out of two of the pathotypes, though I have some reservations about the statistical analyses and need to be reassured that these were carried out appropriately. If this proves to be so, the authors have demonstrated a case of evolution of enhanced in natural populations of *P. triticina*.

This is interesting and useful. However, I have some issues with the context (as well as with the statistical analyses). The authors frame their questions around quantitative (as distinct from qualitative) resistance, and the relation between aggressiveness and quantitative resistance is unclear to me. If aggressiveness or its components are fitness traits for the rust then it is not so surprising that aggressiveness will increase, because more aggressive genotypes should out-infect and out-transmit their less aggressive counterparts. This should be the case on host substrates with or without or with more or less quantitative resistance so I do not understand the connection between natural selection for increased aggressiveness and quantitative resistance of the host. It would be helpful if the authors could clarify this relationship.

→ Aggressiveness (and its components) can be defined as the 'parasitic fitness' of the pathogen (e.g. Shaner et al. 1992 doi.org/10.1146/annurev.py.30.090192.000403), which is a fortiori independent from the nature of the resistance in the host (quantitative vs qualitative). For this reason, we remove mentions of quantitative resistance in our manuscript, including the corresponding section of the Discussion.

Nevertheless, the question of host cultivar specificity of aggressiveness remains an interesting scientific question. If an increase in aggressiveness is dependent on the host genotype, then this specificity could be explained by a difference of quantitative resistance between host genotypes. However, this question is not addressed in our manuscript.

I understand that the phenotypes the authors measured as components of aggressiveness are expressions of, at the same time, the pathogen's ability to exploit the host and the host's degree of defence. However, one or both of these may vary, and the authors need to explain this and how they decide whether they are looking at host traits (quantitative resistance), pathogen traits (aggressiveness) or the interactions of the two. That they tested all pathogen isolates on more than one host may allow them to detect when a particular phenotype varies

among different pathogen genotypes, between or among host cultivars or in interaction, but I could not find formal tests for the interactions and perhaps the design did not allow comparisons of the host cultivars in all cases (but the statistical analyses are presented too superficially for the reader to really tell).

→ *Our study was designed to study components of aggressiveness per se, that is to say the pathogen's ability to exploit the host. We did not measure host traits so we did not measure quantitative resistance, and that is the reason why we decided to remove mentions to quantitative resistance in our manuscript. Though, it is true that components of aggressiveness were measured on several cultivars. It was mostly done in different experimental series (see Tables 1 & 2) that did not allow us to statistically test for interaction between pathogen's genotype and cultivar.*

And for another smaller issue, it is suggested that "aggressiveness" implies "causing more damage to the host", i.e., more aggressive isolates will impact their hosts more. Of course it makes sense that higher infection efficiency, shorter latency period and higher spore production per lesion will cause more host damage, but this is always implicit and never explicitly explained or justified. I think the authors need to qualify this, either by explaining that these traits DO cause more host damage and give us the evidence or demonstration that biomass or yield or seed quality or whatever is decreased more by more aggressive isolates, or by stating that this is an underlying implicit assumption that has not yet been demonstrated.

→ *We added a sentence in the introduction (line 73-75) stating this assumption.*

I provide a number of detailed comments below, in order of their appearance in the manuscript. Some are just details of wording. Some are more substantial.

Abstract:

Line 18 : "… with a particular combination of"

→ *It has been changed in the text.*

Line 20-21 : this is rather vague. What aspect "could not be explained by..."? Do you mean that, given the R genes in the landscape, you would NOT have expected these pathotypes to decline? Were these pathotypes MORE or LESS common than you would expect from the r-profiles of the cultivated varieties? Please be specific and give as much information as possible.

→ *Their domination and also their frequency evolution in the landscape could not be explained only by the Lr genes because of the presence of others compatible pathotypes at very low frequency in the landscape. We have added more information line 20-22 to be more precise.*

Line 25 : **the more** recent genotype was more aggressive than the older one

→ *It has been changed in the text (line 27).*

Line 28-29 : your description of the “neutral” cultivar is not really adequate. If you need to put quotation marks this means that there is a problem with terminology you have not solved. Please try to explain this better. What do you mean by “no selection effect …”?

→ *A neutral cultivar is a variety that did not statistically affect the pathotype frequency in the landscape. Therefore, such a cultivar can be postulated to have no or minor selection effect on the population composition. We changed the text to explain this better in line 30-32.*

Line 29 : For pathotype 106 314 0, the most recent genotype had a shorter latency period

→ *It has been changed in the text line 33.*

Line 33-36 : A gain in aggressiveness allowed the maintenance of a declining pathotype, and even further expansion of that pathotype, in the pathogen population” What does this mean? If the pathotype is expanding how is it also declining?

→ *The pathotype frequency was declining before the apparition of the new more aggressive genotype, after which, the pathotype frequency has stabilised in the landscape and even increase. We changed the text for more clarity line 37-40.*

“providing evidence that virulence alone is not sufficient, aggressiveness also being required for the adaptation of a pathogen to a changing varietal landscape.

Sorry but if "adaptation" means "evolution by natural selection" this makes no sense. What do you mean by "adaptation to a a changing varietal landscape"? Perhaps you mean that "Adaptation to a changing varietal landscape will not only affect/modify virulence but will also lead to changes in aggressiveness"?

Is this what you mean?

→ *Indeed, this is what we meant. Accordingly, we modified the text line 41-44.*

Introduction:

Line 51-54 : You define “aggressiveness” in terms of pathogen damage to the host

→ *Not really; we defined aggressiveness as “the quantitative variation of pathogenicity on a compatible host” (line 60-61), which lead to damage to the crop plant and determines the rate at which a given disease intensity is reached.*

Line 60 : you list the characters used to assess “aggressiveness”. Do these estimate damage to the host ? can you justify this ?

→ *We added a sentence in the introduction (line 70-73) to clarify this point.*

Line 63 : “infection efficiency is calculated as te proportion of spores that cause a new”

→ *It has been changed in the text line 84.*

Line 69 : “the latency period is the length of time between inoculation and first sporulation

→ *We changed the text (line 89-91) to clarify the definition refering to “first sporulation”, using the reference proposed by Madden et al. (2007).*

Line 83-86 : you state that sporulation capacity is dependent on the latency perdio but this does not automatically follow. Of course, interactions with longer latency periods will start to sporulate later/ However, theur sporulation capacity AFTER they have begun to sporulate may be independent of the LP.

→ *Indeed, but we made this point to emphasize that it is important to adapt the measurement of sporulation capacity in order to take into account the variations of latency period and to disregard it.*

Line 93 : what does this "essential" mean ?

→ *It has been changed into 'important driver' line 115.*

Line 101-106 : I have several issues with this paragraph :

You are not describing "population dynamics" here. I think you are describing "which genotypes or pathotypes dominate the pathogen population" or "pathogen population composition"

→ *We changed "pathogen population dynamic" by " evolution of the composition of the pathogen population" (line 124-125)*

Why do you expect quantitative resistance to select for increased aggressiveness? You tell us in Line 47 that qualitative resistance leads to an incompatible reaction, i.e., no infection. Hence it is pretty clear why qualitative resistance should select for the emergence of the corresponding virulence types. However, it is not clear what phenotypic effect quantitative resistance genes that you mention here have on the pathogen. Without this being clear (do quantitative resistance genes reduce infection efficiency or sporulation capacity or increase latency period? Is that why you expect them to exert selection on these traits?) we do not understand the link between quantitative resistance and your measures of aggressiveness.

→ *Indeed, quantitative resistance of the host and aggressiveness of the pathogen are two phenomena that are not necessarily linked. Consequently, we removed mentions of quantitative resistance in our manuscript as it was not the purpose of our study.*

You state that evolution of increased aggressiveness has been observed (but you do not tell us if this was in a context of quantitative resistance so the connection here is not clear at all.

Indeed, there may be selection for increased aggressiveness, IF aggressiveness increases reproductive success or fitness of the parasites. But that can happen in presence or in absence of quantitative resistance so it is not clear what these observations bring to the question about quantitative resistance.

→ *As explained above, we removed mentions of quantitative resistance in our manuscript. Our scientific question concerns the evolution of aggressiveness independently on the presence of quantitative resistance or not.*

Line 119: "Variation in…" but what does this last sentence refer to ? variation over time for what time scale ? what is the time scale of "the complete life history of a pathotype" ? what is the "life history of a pathotype" ?

→ *The time scale is dependent on the life history of each pathotype, and the life history corresponds to the presence of a pathotype in the landscape, from its emergence to its disappearance. "Life history" has been changed to "lifespan", line 145-148.*

Line 123 : can you be more precise here? Perhaps "to determine whether information on aggressiveness allows us to predict changes in pathotype frequency" or something like that ??

→ *It has been precised line 150-153.*

Line 124 : again "the life history of what" ?

→ *"Life history" has been changed to "lifespan", line 145 and 157.*

Line 125-126 : why are these pathotypes "good case studies?" Can you justufy this statement?

→ *We justified these two pathotypes for being good case studies because of their long lifespan and high frequency in the landscape over the 2005-2016 period; line 156-158.*

Line 127-128 : we first characterized isolates of these two pathotypes using microsatellite markers, to quantify their genotypic diversity".

→ *The sentence has been changed in line 160.*

Line 131 : I think you do not need to give the pathotype identification twice in the same parapraph

→ *It has been removed.*

Line 131-132 : please replace "when relevant" by "for pathotype 106 314 0 only, (ii)…"

→ *It has been replaced in line 163-164.*

Methods:

Line 138 : "were" in place of "have been"

→ *It has been replaced in line 170.*

Line 143-144 : how was a signle pustule isolate obatined from a bulk harvest ? do you mean "single spore"? or was one pustule per leaf sampled at the time of collection before the urediniospores from the leaf were bulked ?

→ *No, we really meant "single pustule obtained from a bulk harvest". To obtain a single pustule, susceptible plants are first inoculated with a low spore concentration from the bulk harvest. Then, when pale flecks appear, before sporulation, only one leaf carrying a single lesion is kept (other leaves are cut away) until it becomes a sporulating pustule. This method is described in Goyeau et al. (2006) to which we are refering in our manuscript.*

Line 147-148 : perhaps here you could already tell us the years spanned by these samples and perhaps explain why you chose this short period for 166 317 0 versus the longer period for 106 314 0.

→ *These precisions were added line 179-185.*

It appears that you assume that the bulk sample from the leaf share the pathotype of the sampled pustule from the leaf. Is this assumption reasonable? Did you ever find more than one pathotype per leaf ?

→ *No, we do not make this assumption, and we observed in several cases that more than one pathotype is present in the same bulk harvest. This phenomenon could explain the "other genotypes" seen on Figures 4 and 5. However, as we did not pathotype isolates with a different genotype, we cannot confirm this hypothesis. We changed the text in the Results line 360-363 and line 382-384 to make this point clear. This point has also been raised in the last part of the discussion line 644-653.*

Line 161-164: Did you genotype "one isolate" of the 44 or "one isolate per pathotype" of the 44?

→ *We genotyped all of the 44 isolates. It has been clarified in the text, line 198-201.*

You use the terms "pustule" and "**uredinium**". Are these equivalent ? if that the case please choose one and use it throughout.

→ *Indeed, both terms are equivalent. We decided to use uredinium/uredinia throughout the manuscript.*

Line 215-218: "three components of aggressiveness, latency period, infection efficiency and sporulation capacity, were assessed for…"

→ *It has been changed line 252-253.*

Line 219: Is it possible to tell us something about Apache's resistance profile ?

→ *Apache carries the two Lr genes Lr13 and Lr37, specified in Table S2. Pathotypes 106 314 0 and 166 317 0 are virulent on these two Lr genes.*

Line 227-229: "in series 1 and 2, we tested whether the two genotypes of each pathotype differed in aggressiveness on…"

→ *It has been changed line 266-268.*

Line 231-236: Series 3, 4 and 5 tested the difference in aggressiveness between the two genotypes 106 314 0-G1 and 106 314 0-G2 on some or all of the wheat cultivars; Aubusson, Premio, Michigan Amber, Sankara, Expert and Bermude — all of which, except Michigan Amber, were among the 35 most frequently grown cultivars in the French landscape during the 2006-2016 period.

→ *It has been changed line 272-277.*

Line 239-240: this statement is not consistent with Table S3 where I find

Isolate	Pathotype	Genotype	Series	Cultivar sampled	Cultivars tested
BT12M119	106 314 0	106 314 0-G1	1+4	Apache	Aubusson, Premio, Michigan Amber
BT12M033	106 314 0	106 314 0-G2	1+4	Apache	Aubusson, Premio, Michigan Amber

The table suggests that either series 1 or series 4 tested on Aubusson and Premio. However, these two strains should also have been tested on Apache, no?

→ *Indeed, these two strains were tested on Apache. We added this cultivar in Table S3.*

Line 242-243: here or somewhere (perhaps in the table?) you should state on what dates these replicated and series were carried out

→ *The dates were added for each serie, line 266-275.*

Line 252: "under binocular magnifier" should be "under a dissecting microscope"

→ *It has been changed line 298.*

Statistical analyses:

Your presentation of the statistical analyses are not really sufficient though this problem is much worse in the results. I cannot tell how your tests were performed and how you dealt with the hierarchical nature of your data (several observations per isolate, several isolates per genotype) for the non-parametric analyses

All statistics, be they parametric or non-parametric, assume that each datum has the same degree of independence from all other data. This is not the case in your design and must, therefore, be analysed with a hierarchical model. Alternatively, you need to take the mean or median and analyse that. You certainly MAY NOT consider all data as though they are independent.

→ *Series 1 and 2 were analyzed independently; Series 3 and 4 were analyzed together but for each cultivar independently; Serie 5 was analyzed for each cultivar independently. For more clarity, we are now providing two additional supplementary tables (Table S5 and S6) with all p-values and degrees of independence used in our statistical analyses.*

Also, we agree with Reviewer 2 that all data do not have the same degree of independence, so a hierarchical model has to be used. We confirm that a hierarchical model has been used to perform these ANOVA analyses, including Isolates within Genotypes: "the trait I (isolate) is nested within the trait G (genotype)". This was clarified in the text line 329.

You appear to not test the interaction between cultivar and genotype. Why?

For series 3, 4 and 5 maybe you can test the three cultivars common to all (Aubusson, Premio, Michigan Amber) using Model 1. Might this give additional information?

→ *We did not test the interaction between cultivar and genotype because it was not the purpose of the study, which was not designed for that but rather to test the hypothesis of a variation (expressing an evolution) in aggressiveness within the same pathotype. Moreover, different cultivars were mostly used in different experimental series, with different genotypes for each pathotype, so the data were treated for each variety separately, except for Series 1 and 2 where both Apache and Michigan were used simultaneously and with the same isolates.*

Line 287-288: Why do you write “to analyze the effect of genotype on the aggressiveness components.” You also test the cultivar effect in some cases

→ We test the cultivar effect only in Series 1 and 2, where both cultivars Apache and Michigan were used simultaneously and with the same isolates. We added a sentence to clarify this point line 336-338.

Results:

Line 295-296: Your description of the dynamics is not complete. Either just refer to the Figure, i.e., “The dynamics of the two pathotypes 106 314 0 and 106 314 0 in the French P. triticina population are shown on Figure 3.” Or describe more completely, i.e., “The frequency of pathotype 106 314 0 in the French P. triticina population increased from 30% in 2006 to 51% in 2009 and decreased back to about 30% in 2011 (Figure 3). After a plateau at 30-33% from 2011 to 2014, the frequency of this pathotype decreased strongly, to less than 1% in 2018.

→ We described Figure 3 more precisely, line 345-349 and 369-373.

Line 299: this should be figure 4

→ Indeed, it has been changed.

Line 301-307: I think this can be summarized as follows: “Genotype 106 314 0 was the rarer of the two until 2012 after which it replaced genotype 106 314 0-G2. (Figure 4)

→ We prefer not to change this sentence and be more precise in the way we describe the temporal change in the frequencies as this is an important aspect of the study.

Line 307-310: since you told us that the cumulative frequency of G1 and G2 varied between 40 and 65% we already know that there were others genotypes that represented the remaining 35-60%. I think you can delete this.

→ We deleted the sentence line 366 and line 384-385.

Line 313: “Its dynamics … showed two peaks …” Do you tell us how many samples were examined each year to generate these curves? How much confidence can we have in the first peak? Maybe it is just a sampling blip due to small sample sizes per year. I think here you also over-describe your results. Just say that the dynamics are shown in the figure. That is really enough

→ A significant number of isolates has been sampled each year, as you can see in the following Table (data available in Fontyn et al., 2022: https://doi.org/10.1111/ppa.13599; referenced in our manuscript). Thus, we are confident about the evolution of the pathotype frequency in the landscape and we think it is important to describe its dynamic in this Results section. We specified the total number of each pathotype in a sentence added to the legend of Figure 3 to emphasize the robustness of the analysis that these data allow.

	Year of sampling										
Pathotype	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
106 314 0	87	109	104	99	87	107	133	102	112	73	12
166 317 0	0	6	12	15	38	68	131	62	82	56	68
others	219	303	122	78	92	156	147	131	183	287	165
Total	306	418	238	192	217	331	411	295	377	416	245

Please trim and reduce the paragraph on lines 312-324.

→ We believe that it is important to keep in our manuscript the description of the frequency of pathotype 166 317 0, and its genotypes -G1 and -G2, in the landscape during the period considered.

I suggest: "Pathotype 166 317 0 was first detected in 2007 at a very low frequency (less than 2%; Figure 3). Two genotypes of this pathotype, differing by only one of the 19 SSR loci used (RB8; Table 3), were observed from 2013-2016 period, during which the initially dominant 166 317 0-G1 genotype was almost replaced by the 166 317 0-G2 genotype (Figure 5). Over the four-year period other genotypes accounted for 30% to 46% of this pathotype."

→ In the suggested sentence, the evolution of the pathotype frequency in the landscape does not appear, which is important to keep in order to compare later with the evolution of aggressiveness. This is why we would like to keep details about the frequency evolution of each pathotype.

Line 326-337: We need to see the analyses to be able to judge the reliability of these results. There is a problem of experimental design if you carry out a non-parametric analysis, I think, because I do not understand how your degrees of freedom are calculated for the nested analysis. You must perform a nested analysis or correct for non-independence of data somehow. I think you have several observations per isolate and several isolated per genotype so this is clearly a hierarchical structure that must be analysed appropriately. Without seeing the calculations of the statistics with their dfs the reader cannot know that you have analysed your data correctly.

Without seeing your ANOVA tables and understanding how you dealt with data non-independence I cannot judge whether you found differences in aggressiveness or not

→ As explained above in the Statistical Analyses, we clarified the text in the Materials and Methods page 14, and we added supplementary Tables S5 and S6 with details on the ANOVA performed.

Discussion:

The discussion is very long. Also it sometimes goes beyond where the data comfortable allow. The authors should soberly look at what they really have evidence for and stick with that.

I think there may be some over-interpretation of the results.

For pathotype 106 314 0 you find very little evidence for an increase in aggressiveness. Nonetheless you interpret the reduction in rate of decline in frequency of the pathotype between 2011 and 2014 as being due to a new more aggressive genotype arising. First, the genotype that increased in frequency was not much more aggressive.

→ We found that within both pathotypes 106 314 0 and 166 317 0, the most recent dominant genotype was more aggressive than the older one. This was true for both pathotypes, even so it was only the case for one of the three components of aggressiveness (the latency period) concerning 106 314 0. It shows that a gain in aggressiveness may depend

on the component(s) considered. This evidence is statistically significant on several cultivars (5/6; Table 4) so cannot be considered "little" as suggested by Reviewer 2. Moreover, small effects, identified in a controlled environment during one reproductive cycle of the disease, may result in a strong impact on the pathotype frequency in the landscape due to the polycyclic nature of rust epidemics.

Second, it is not clear to me whether you genotyped all the isolates that are featured in Figure 3. Can you tell us the genotype frequencies of the two strains on that curve? If you have those data please plot them. You can make pie diagrams for each point to show the genotype frequencies for each observation. Without this the arguments are not that convincing. I really need more information to connect what I see on Figs 4 and 5 with the curves on Fig 3.

→ *The isolates from the survey, purified from spore bulks, have been pathotyped (featured on Figure 3) and then discarded. Unfortunately, these purified isolates have not been genotyped before being discarded, so their genotype frequencies are not available. For this reason, we purified new isolates from 401 spore bulks in which pathotypes 106 314 0 (n = 286) or 166 317 0 (n = 115) were previously identified. These 401 purified isolates were genotyped (featured on Figures 4 and 5) but were not pathotyped. Consequently, isolates featured on Figure 3 are different but related (from selected spore bulks) to isolates featured on Figures 4 and 5. The whole procedure is described in the first section of the Materials & Methods and on Figure 1. The total number of isolates genotyped are given in Figure 1, but this number is now added in the legend of Figures 4 and 5.*

Line 468: What is "mid-term"?

→ *The sentence has been changed line 534.*

Line 478-479: Do we really know that the change in frequency of pathotypes is driven by changes in resistance gene frequencies? Do we know what resistance genes are in the landscape? If we do, we should be able to predict which pathotypes will increase or decrease. Can we do that?

→ *We know that changes in pathotype frequencies is driven by changes in resistance frequencies, as shown by Fontyn et al. (2022: https://doi.org/10.1111/ppa.13599). The introduction of a new Lr gene in the landscape is followed by the emergence of the corresponding virulence, and also an increase of frequency of a Lr gene is followed by an increase of the corresponding virulence. This dynamic has been described in great detail in our INRAE research group. Through our collaborations with wheat breeding companies and French extension services we also now, in general, which Lr genes are present in wheat cultivars and are mostly found in the landscape. However, as established by Fontyn et al. (2022), these information on Lr gene frequencies did not explain alone the domination of the two pathotypes used in this study. This is actually the rational of our hypothesis on the role played by aggressiveness in pathotype frequencies in the landscape, that we tested in the current study.*

Line 551: You qualify your G1 genotypes as "oldest" but you do not know their history, I think. Similarly the G2 may not be younger, necessarily. Perhaps these genotypes emerged in the past and happened to be at low frequency at the beginning of your sampling period. For 166 317 0 you cannot really know which one is older though they appear to be closely related.

→ We believe that the genotypes G1 are older than the genotypes G2, as for both pathotypes genotypes G2 were detected for the first time later than genotypes G1. We agree that it remains an assumption, even if it is the most likely.

Line 556: Differing at several loci does not mean that 106 314 0 has "more genetic diversity". Here what you can say is that the two genotypes are more differentiated. You would need to compare the number of different genotypes in similarly sized samples to compare genetic diversity.

→ This sentence has been changed with "are more differentiated", line 622-623.

Line 557-559: The relationship between microsatellite, presumably neutral mutations and phenotypically relevant mutations is not very straightforward. That the two genotypes of 106 314 0 are more differentiated for these neutral markers does not necessarily mean that they are more differentiated for functional mutations. In fact, you found LESS phenotypic variation for your aggressiveness traits.

→ We agree with Reviewer 2. That is the reason why we stated two different hypotheses to explain this differentiation between genotypes of pathotype 106 314 0. It remains only hypotheses to be tested in the future, using genome-wide genotyping approaches for example.

Line 603-604: what does "Conversely, identical genotypes may differ in one or several virulences" mean ?

→ It actually corresponds to the previous comment: differentiation (or not) for neutral markers is different and partly independent from differentiation (or not) for functional mutations. In other words, two identical microsatellite genotypes can have different virulence profiles. We added a sentence to summarize this point line 671-672.

I am not sure that "pathogenotype" will be a useful term.

→ We found the term "pathogenotype" useful to rapidly distinguish between groups of isolates having a unique association between one pathotype and one genotype. We would like to use this term in future works and manuscripts concerning population surveys of leaf rust and other diseases. As this term is not wrong per se, we propose to keep it at the end of our discussion and to let the scientific community decide of its usefulness in the future.