Even though authors responded to most of my remarks and comments, I'm still not fully satisfied with some of their answers. In addition, I compared the new manuscript with the old one and realized that not all changes being made by them could be followed using the track-changes functionality in Word. Since several changes were made, I decided to check the whole manuscript again and find some additional minor points that should be assessed by the authors.

Respecting the old remarks that were not properly assessed by the authors, I'm including here my original comment [Original remark], the response of the authors to my remark [Response] and my new remark [New remark] to these answers:

[Original Remark] Page 12, line 171: Please explain the equation by describing its different components.

[Response] We apologize the mistake for citation and revised the section in the text (Line660-663, page24).

$p_i$ and $p_j$ were the frequencies of the number of *i* and *j* alleles respectively, and *n* was the number of alleles (Botstein et al.,2008).

Following Table gives the mating types, their frequencies, and the probability of informativeness of an offspring for a marker locus, $A_i$, with n alleles and frequencies $p_i$.

| Genotype of affected parent | Genotype of unaffected mate | Probability offspring informative | Frequency of mating |
|---|---|---|---|
| $A_i$ $A_j$.................. ........ ........ ...... | $A_k A_j$ | 1.0 | $2p_i\, p_j\, (I - p_i - p_j)^2$ |
| $A_i$ $A_j$.................. ........ ........ ...... | $A_k A_i$ or $A_k A_j$ | 1.0 | $2p_i\, p_j\, (I - p_i - p_j)\ (2\, p_i + 2\, p_j)$ |
| $A_i$ $A_j$.................. ........ ........ ...... | All homozygotes | 1.0 | $2p_i\, p_j \sum_{i=1}^{n} p_i^2$ |
| $A_i$ $A_j$.................. ........ ........ ...... | $A_i A_j$ | .5 | $2p_i\, p_j$ X $2p_i\, p_j$ |
| All homozygotes | All genotypes | 0 | $p_i^2$ |

One can evaluate each marker locus for its polymorphism information content (PIC) by summing in table 1 the mating frequencies multiplied by the probability that an offspring will be informative. Under our assumptions, the expected value of PIC can be calculated as

$$\mathrm{PIC} = 1 - \sum_{i=1}^{n} P_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2P_i^2 P_j^2$$

[New remark] The explanation of the equation is clear for me here, but not for the readers. Please summarize and explain the components of the equation in the text. You cannot write an equation without explaining its components.

[Original remark] Page 12, line 190: 0.0008 corresponds to an Euclidean distance? Please specify.

[Response] Yes, It corresponds to the Euclidean distance.

[New remark] For me, this is clear, but not for the readers. Please specify this in the text.

[Original remark] Page 13, lines 206-207: How did you calculated the R-squared? Was this base on the mixed model or using a simple linear regression model? Please specify. In addition, R-square is the coefficient of determination, not a correlation coefficient.

[Response] We thank the reviewer to point this issue. The R-square value was from the mixed model, and calculated in the TASSEL program. Yes, the R-square at this place is the coefficient of determination.

[New remark] The way the R-squared was calculated has to be mentioned in the text. In addition, you do not replaced "correlation coefficient" with "determination coefficient" in the text. Please correct.

[Original remark] 2.Page 20, lines 370-371: "Associated traits were characterized under two different environments, suggesting that the significantly associated SNPs were genuine and stable". I do not agree with this idea. A dataset with two environments is not robust-enough to make such a conclusion. Were the two environments similar or contrasting? How was the correlation of phenotypic values among them? Please clarify and improve.

[Response] We appreciate reviewer's comment. In this revision, we supplemented the part of content about the description of two different environment and added the analysis of the correlation coefficient of the same trait between two environments. Which were shown in Table 4 and the text (Materials and Methods section, line169-170, page 7; result section, line308-309, page 12; discussions section, line411-412, page 15).

[New remark] You are still arguing that "The same trait was significantly correlated between two different environments, suggesting that there are little false positive in the significantly associated SNPs". I do not agree with this. If you found associations in both environments without a change in direction of allelic effects, this only means that associations are relatively environmentally stable, but this does not mean that they are not false positives. False positives can arise because of many reasons during association mapping. If you do not find any convincing argument for this idea, please eliminate from the manuscript. Please check the WHOLE manuscript for every section where this kind of comment appears.

[Original remark] Page 20, lines 373-375: "Leaf size is associated with local adaptation, and it can actually decrease the false rate by removing accessions with extreme genetic diversity and phenotypic expression". Please explain in more detail how this decrease in false rate happens.

[Response] we removed these accessions that were seriously affected the growth by diseases, insect pests and marginal effects to decrease the false rate.

[New remark] What you are mentioning here has to be mentioned in the manuscript.

[Original remark] Page 5, line 1: I suggest to use "development of core-collections" instead of "development of the core-collection". Otherwise sounds to me like this is and will be always the one and only tea core-collection developed.

[Response] We thank the reviewer for your suggestions and have revised the title.

[New remark] This was not properly corrected. It should be "core-collections".

[Original remark] Page 8, line 66: Please check and correct the character size of "the linkage disequilibrium (LD)", because it is not matching the rest of the text.

[Response] We thank the reviewer to point this issue and have checked and corrected in the introduction section.

[New remark] This was not properly corrected.

[Original remark] Page 8, line 68: Please replace "; other advantages" with ". Other advantages".

[Response] We have replaced.

[New remark] This was not properly corrected, there is still a semicolon ";" in between that should be replaced by a period ".".

[Original remark] Page 8, lines 69-70: Please replace “and less time-consuming for developing” with “and providing a less time-consuming approach for developing”.

[Response] We have replaced.

[New remark] You eliminated too much text. It should be “and providing a less time-consuming approach for developing the mapping population”.

[Original remark] Page 13, line 205: Please try to stick to one single format number. For instance, write “30,282” instead of “30282”. Check the format consistency of numbers greater than 1,000 throughout the whole manuscript, figure captions, supplementary material, etc.

[Response] We thank the reviewer to point this issue and have checked and revised.

[New remark] I found still several inconsistencies throughout the whole manuscript. See for instance the description of the supplementary tables. Please check again the whole manuscript and correct.

[Original remark] Page 15, line 266: Please replace “of the whole tea accessions, separately” with “of those of the whole population, respectively”.

[Response] We thank the reviewer to point this issue and have replaced.

[New remark] There is still an error. Please check and replace “and 98.25% of the those of the whole” with “and 98.25% of those of the whole”.

[Original remark] Page 15, line 272: Please replace “1). The proportion of samples number removed from per group” with “1), respectively. The proportion of samples number removed per group”.

[Response] We thank the reviewer to point this issue and have removed.

[New remark] There is still an error. Please check, after mentioning “(Fig.1, Table 1)” should come immediately a “, respectively” because you enumerated different values for the different core-collections.

[Original remark] Page 17, line 301 to page 18, line 302: Please replace “; a negative effect of major alleles of -1.31 and a positive effect of minor alleles of 0.44 were detected” with “. This SNP has a negative major allele effect of -1.31 and a positive minor allele effect of 0.44”. In addition, please specify the trait you are referring to.

[Response] We thank the reviewer to point this issue and have replaced and revised.

[New remark] There are still errors. Please check, it should be “… phenotypic variation. This SNP has a negative major allele effect of -1.31 and a positive minor allele effect of 0.44 for MLW. Six loci, located …”.

[Original remark] Page 19, lines 348-349: Please replace “distance-based methods to propose the first core collection and the mini-core collection from tea origin center, the Guizhou plateau, by GBS, accounting” with “distance-based methods applied on GBS data to propose the first core collection and the mini-core collection from the tea origin center, the Guizhou plateau, and accounting”.

[Response] We thank the reviewer to point this issue and have replaced.

[New remark] Please double check this, this sentence looks exactly the same as it did before.

In addition I found the following details that should be assessed. The position in the modified manuscript are indicated:

General comment: There are still some text formatting issues in the manuscript and you must double check them before sending the new version of your manuscript. It is really time consuming to enumerate all of them, but, for instance, prepositions starting with capital letters in the middle of the sentences or following semicolons or commas, in addition to sections with two blank spaces together (for instance in the reference section), or a period followed directly by a letter (with no space in between), or an abbreviated word presented several times in its complete form in the manuscript or introduced several times like, for example, marker-assisted breeding (MAB) or neighbor-joining (NJ). These are only some examples and could be more.

Page 2, line 27: It must be “set were very well” and not “set very well”.

Page 4, line 85: There are two dots “..” following “Frankel” and “Brown”. This has to be eliminated. Please check for this kind of formatting error throughout the whole text.

Page 4, line 93: It must be “sunflower” and not “suflower”.

Page 5, lines 115-116: it must be “…years old will be called “ancient landraces”, and the tea garden landraces will be named “modern landraces” (Table. S1). The ancient …“.

Page 7, line 169: it must be “mature leaves width” and not “mature leaves with”.

Page 7, line 170: for correspondence of the text, it must be “MLL and MLW” and not “MLW and MLL”.

Page 8, line 185: it should be “relationships of accessions” instead of “relationships of among different accessions”.

Page 9, lines 221 to 222: it should be “to cut down the rate of false positives” instead of “to cut down the false positive among genotypes”.

Page 9, lines 228 to 229: it should be “and the relevant information” instead of “and the information”.

Page 9, lines 229: After the citation there is a period “.” missing.

Page 9, lines 240-241: It should be “Based on the NJ cluster analysis on the genetic distance matrix, the 415 …” instead of “Based on the genetic distance matrix yielded from the NJ cluster analysis, the 415 …”.

Page 10, lines 245 to 257: I do not understand the percentages included in brackets. Shouldn’t the percentages within a group summed to 100%? Please clarify in the manuscript.

Page 11, line 291: “increased remarkably” is quite exaggerated for a ~1% increment. I suggest to use “slightly increased” instead.

Page 12, line 322: it should be “significant associations” instead of “significant association”.

Page 12, lines 327 to 328: it should be “phenotypic variation. This SNP has a negative major allele effect of -1.31 and a positive minor allele effect of 0.44 for MLW.”.

Page 13, line 340: it should be “significantly associated” and not “significantly association”.

Page 14, line 363: the word “population” is followed by “… It must be only one “.”.

Page 14, lines 369 and 376: “marker-assisted breeding” should be abbreviated for consistency of the text.

Page 14, lines 377 to 378: it should be “methods to develop”.

Page 14, line 383: please replace “almost were same as the whole tea accessions , which” with “were almost the same as for the whole population, which”.

Page 15, line 397: it must be “was equally appropriate” instead of “was same appropriate” .

Page 15, line 400: it must be “selected in this study is easy” instead of “selected in this study are easy”.

Page 15, line 412: it must be “significantly correlated” instead of “significant correlated”.

Page 15, line 419: it must be “GLM” instead of “GML”.

Page 15, line 420: it must be “which reduce the rate of false” instead of “which reduce rate of false”.

Page 16, line 435: after the second comma “,” should go “which was consistent”.

Page 16, line 436 to 438: Please check for punctuation problems. It should be “three characters (Table 4).”.

Page 18, line 496: it should be “frequencies” instead of “frequencies frequencies”.

Page 18, line 505: it should be “analyzed” instead of “analysed”.