------------------------------------------------------------------------------------------------------------------------Associate editor comments 15.06.20  
------------------------------------------------------------------------------------------------------------------------  
Both reviewers have provided a number of comments that should be considered by the authors. Especially, Reviewer 1 asks for a detailed theoretical part, which is completely missing. I strongly agree with this request. Furthermore, this reviewer suggests to remove the pig breeding example from the paper. The application demonstrates some properties of the R package (see also comments from Reviewer 2) and therefore, I encourage the authors not to remove this part, but to describe it in a better way (see suggestions from Reviewer 2).

We thank the editor for the comments and directions. We adjusted the manuscript to account for reviewers’ comments. As suggested, we did not remove the multi-tier breeding example, but we did add a brief description of the method and substantially changed the structure of the results to make them more comprehensive for the readers.

------------------------------------------------------------------------------------------------------------------------  
Reviewer #1

------------------------------------------------------------------------------------------------------------------------  
This is a poorly written paper on a very interesting R package. The aim of the manuscript should be to introduce readers to the package and to have the role of reference for people who apply the package. This requires, first of all, a theory part, which is completely missing in the manuscript. The package is applied to a pig breeding example. The pig breeding example is not part of the package and does not contribute to the understanding of the package. It should therefore be removed from the paper. It may be suitable for a separate paper that focusses on pig breeding.

We thank the reviewer for the comments. We addressed the comments and suggestions in the revised manuscript. We added a brief theoretical background of the method. We did not remove the pig breeding example from the manuscript since we believe, it provides a good example of the type of questions that could be tackled with the software, how to interpret the results, and how the results can help to better understand the dynamic of genetic gain in breeding programmes. The editor agrees with keeping the example.

My comments in detail are:

------------------------------------  
Line 65: Could all these studies have been performed with the AlphaPart package? Please state it here.  
  
Thank you for the comment. No, these studies were not run with AlphaPart, but with a set of scripts that we consolidated into the AlphaPart package. We reworded the sentence to make it clearer (lines 76-78): “*However, these studies used bespoke implementations of the partitioning method, for which no open-source software exists.”*

------------------------------------

Line 69: A theory part is missing. A theory part should define the quantities that are computed by using population genetic and quantitative genetic terms and should provide the formulas for computing them. In particular, the theory part should clearly define the components into which the breeding values are partitioned. It should define what a path is, and how the parts are assigned to paths.

We thank the reviewer for this remark. We included the theory part in the revised manuscript (lines 55-70). However, we provide only a brief description, since the method was not developed within the scope of this research and is described in detail in the paper of García‑Cortés *et al.*, 2008.

Added text: “*In summary, the method uses pedigree to first partition the breeding values into parent average and Mendelian sampling term: ai = ½as + ½ad + wi [4], where ai, as, and ad are individual’s, sire’s, and dam’s breeding values, and wi is individual’s Mendelian sampling term. The method next allocates Mendelian sampling terms to different “paths”. For example, assume a small trio pedigree with two parents and a female progeny. Specifying gender as the path variable, we can write ai = ½as + ½ad + wi = (½ad + wi) + ½as  = ai,f + aim, where the first term ai,f denotes the female partition and the second term ai,m denotes the male partition. Alternatively, assume that the sire is imported, then an interesting path specification is that of domestic versus imported, which in the example gives the same partitioning as for gender. In general, we can write a vector of breeding values as a linear combination of Mendelian sampling terms of individuals and their ancestors* ***a*** *=* ***Tw****, where* ***T*** *is a triangular matrix of expected gene flow between ancestors and individuals [4,5]. The method of García-Cortés et al. [3] uses a path variable to partition the gene flow matrix into path specific gene flow matrices* ***T*** *=* ***T****1 +* ***T****2 + … +* ***T****p and with this partitions breeding values by paths* ***a*** *= (****T****1 +* ***T****2 + … +* ***T****p)****w*** *=* ***a****1 +* ***a****2 + … +* ***a****p. Summarising these partitions marginally or conditionally on other variables (such as year of birth) is a powerful way to analyse sources of genetic gain.”*

------------------------------------  
Line 72: This section should be expanded, but only example data sets should be used throughout the paper that are included in the package. The purposes of the data sets are only to explain the functionality of the package and the theory behind it, so the data sets could be small.

Thank you for the suggestion. We now added the example inputs and outputs of the functions.  
------------------------------------  
Line 82: Please append parentheses after function names throughout the paper, i.e. replace AlphaPart by AlphaPart().

We added the parentheses to all function names throughout the paper.  
------------------------------------  
Line 84: The first lines of data frame AlphaPart.ped should be shown and explained in the text.

We additionally explained the variables of the pedigree (lines 93-96) and added an example of the dataset to the manuscript (lines 97-102).

*“The package includes an example dataset AlphaPart.ped. The dataset includes a four‑generation pedigree (individual IId, father FId, and mother MId) and information about the generation (gen), country, gender, and breeding values for trait 1 (bv1) and trait 2 (bv2).*

*> head(AlphaPart.ped, n = 3)*

*|IId |FId |MId | gen|country |gender | bv1| bv2|*

*|:---|:---|:---|---:|:--------|:------|---:|---:|*

*|A | | | 1|domestic |F | 100| 88|*

*|B | | | 1|import |M | 105| 110|*

*|C |B |A | 2|domestic |F | 104| 100|*

*“*

------------------------------------

Line 87: The content of the variable part should be shown (or its structure) and explained.

We added the example output and described the structure. Lines 123-136: *“The output of the AlphaPart() function is an object of either AlphaPa**rt or summaryAlphaPart class. The AlphaPart class is a list with an info element and an element for each of the partitioned traits. The info element is a list with information on the path variable (path), number of paths (nP), names of paths (lP), number of traits (nT), names of traits (lT), and putative warnings (warn). The trait element bv1 is a table (shown below). It includes the information from the original dataset as well parent average (bv1\_pa), Mendelian sampling term (bv1\_w), and breeding value partitions (bv1\_domestic and bv1\_import).*

*> head(part$bv1, n = 3)*

*|IId |FId |MId | gen|country |gender | bv2| bv1| bv1\_pa| bv1\_w| bv1\_domestic| bv1\_import|*

*|:---|:---|:---|---:|:--------|:------|---:|---:|------:|-----:|------------:|----------:|*

*|A | | | 1|domestic |F | 88| 100| 0.0| 100.0| 100.0| 0.0|*

*|B | | | 1|import |M | 110| 105| 0.0| 105.0| 0.0| 105.0|*

*|C |B |A | 2|domestic |F | 100| 104| 102.5| 1.5| 51.5| 52.5|“*

------------------------------------  
Line 91: Please try to avoid "we" throughout the paper. For example, the sentence "we recurse the pedigree from…" could be replaced by "function AlphaPart() recurses the pedigree from…"

Thank you for this comment. We reworded the sentences where appropriate.  
------------------------------------  
Line 93: How is the term "path" defined? How are the Mendelian sampling terms and the parent averages assigned to the paths?

“Path” variable contains groups to which we allocate Mendelian sampling terms, e.g. males-females, different countries, different AI centres, etc. We described the method in the Background (lines 55-70) using the “path” term.

------------------------------------  
Line 95: The posterior distribution of what?

Thank you for the comment. If the users estimate the breeding values with Bayesian methods, they acquire a sample from the posterior distribution. This option allows them to partition individual samples to obtain the information about the uncertainty of the partition. We clarified this in lines 117-118: *“posterior distribution of breeding values”.*

------------------------------------  
Line 97: What do you mean with "on the fly"? What is the alternative?

AlphaPart package includes two C++ executables. One of them only partitions the breeding values and outputs the partitions for each individual. The other one partitions the breeding values and tallies the partitions by group directly without storing the individual values, which can reduce memory use considerably for large analyses. We changed the manuscript (lines 119-122): *“The function can also simultaneously partition and summarize path contributions by a grouping variable, which is a useful computational speed-up for huge pedigrees. Alternatively, we subsequently use summary() function to summarize the partitions.”*

------------------------------------  
Line 103: Please show a part of the result here and explain the result. In general, I would suggest the following: First introduce and explain a function and its general use, then give an example with R code and show a part (or a summary) of the result, and then explain what the function was doing in this special case.

We added the example output and explained the results.

Lines 145-155: “*The output of the summary.AlphaPart() function is an object of summaryAlphaPart class, which is a list with an info element and an element of summary for each of the partitioned traits. The summary contains the grouping variable levels (gen), number of individuals per level (N), and output of the summary function applied to the breeding values and its partitions (Sum, domestic, and import).*

*> print(sumPartByGen)*

*| gen| N| Sum| domestic| import|*

*|---:|--:|--------:|--------:|------:|*

*| 1| 2| 102.5000| 50.00000| 52.500|*

*| 2| 3| 104.6667| 17.16667| 87.500|*

*| 3| 2| 107.0000| 13.37500| 93.625|*

*| 4| 1| 109.0000| 68.87500| 40.125|*

“  
------------------------------------  
Line 109: Please provide a figure with the plot.

We added the plot as Figure 1.  
------------------------------------  
Line 118: The pig breeding example should not be part of this paper because it distracts the reader from the main issue, which is to teach the reader to handle the package and to understand the theoretical background behind it.

We appreciate reviewer’s opinion. In line with editor’s and second reviewer’s comments we decided to keep this example, since it helps to understand the usefulness of the package and interpretation of results in more complex breeding programmes.

------------------------------------------------------------------------------------------------------------------------  
Reviewer #2  
------------------------------------------------------------------------------------------------------------------------  
This is a good paper, and it is great to see the software in the public domain in a readily assessable form. There are a couple of areas where the paper should be improved.

We thank the reviewer for the comment and the recognition.  
------------------------------------  
First, the supposed pig breeding structure used to demonstrate the method is extremely simplistic and generalised. When I read this manuscript, I was expecting to get some deep insights into breeding design for a pig program. However, because of the simplicity of the simulation structure (one cycle of selection and generation per year), it is misleading. I suggest the wording be changed to acknowledge that the example is a stylised multi tiered breeding program, and it might be better to change years, to generations.

Thank you for your comment. We changed “*pig breeding example*” to “*stylized multi‑tier breeding programme*” throughout the manuscript. We also changed “*years*” to “*generations*” throughout the manuscript.

------------------------------------  
Second, the results are quite difficult to interpret. In particular, the comments about multiplier versus nucleus comparisons. My understanding is that the partitioning method partitions genetic trends in merit of animals born, into cumulative contributions to the aggregate trend through selection within defined subsets of candidates.

We are not sure if we understand the comment correctly. The method partitions the parent average and Mendelian sampling terms according to the path variable. In the example we defined the path variable to a combination of nucleus and multiplier, and male and female gender. This enables us to quantify how the selection in two tiers and two genders contribute to genetic trend in the nucleus and the multiplier. Obviously, all genetic gain in the closed nucleus is due to the nucleus selection. Surprisingly, genetic gain in the multiplier is not only due to nucleus (as commonly believed), but also due to multiplier selection (for traits available). We adjusted the manuscript to clarify that we measured the genetic gain separately in the nucleus and the multiplier. We also clarified, which genetic gain we are measuring, partitioning, and describing in a certain part of the manuscript.

------------------------------------

When comparing nucleus versus multiplier animals, what are the comparisons?

We first compared the genetic gain in the nucleus and multiplier to show that multiplier achieved higher genetic gain (and had animals with higher breeding values), which is not expected in a multi‑tier breeding programme. Next we compared the contribution of nucleus and multiplier male and females selection to the genetic trend of the nucleus and multiplier separately. We now added a paragraph describing the structure and flow of the results and adjusted the results accordingly.   
Lines 223-226: *“For each scenario we first describe the distribution of true breeding values in the nucleus and multiplier in generation 40 of one replicate. Next we explain the sources of the observations by partitioning the nucleus and multiplier genetic trend and averaging the results across ten replicates.”*

------------------------------------

In Figures 1 & 3, the distributions  seem to be always about the same shape and width, so what is the value of showing them? I think that some additional effort to better explain the meaning and interpretation of the results would help.

Thank you for this comment. The purpose of these figures was to show both the mean and the spread of the breeding values of the nucleus and multiplier animals of a specific generation. We wanted to show, that i) multiplier has a higher mean than the nucleus for some traits; ii) multiplier can produce animals with higher breeding values than the nucleus, which is unusual. A small change in mean might not mean much, but this can indeed lead to better multiplier individuals than in nucleus (for the measured traits), which has been observed in some real-life cases but could not be explained. We now simplified these figures to include only one generation and only the distribution of the breeding values and not the partitions.

------------------------------------  
Line 34 - the term "the multiplier achieved higher breeding values" is a little confusing.

Thank you for the comment. We meant to say that we observe animals with higher breeding values in the multiplier than in the nucleus. However, the abstract has been reworded and does not include this sentence anymore.

------------------------------------

While it is hard to communicate necessary details in the summary, could this be reworded? Maybe the word "achieved" could be changed to "maintained", maybe "the multiplier path maintained higher average breeding values than…."? Achieved implies some sort of end point, but breeding programs are continuous….

To acknowledge the continuity of the breeding programme, we now replaced the “*final*” genetic gain with genetic gain “*in generation 40*” (which is the last year of comparison). We also replaced “*achieved*” with “*had*” where appropriate.

------------------------------------

Or maybe you mean that the multiplier contributed a higher proportion of the genetic progress in these traits? As per my second point above, I struggled with this through the whole manuscript.

Thank you for the comment. We now clarified in the manuscript that first we compare the genetic gain in the nucleus and multiplier and then move onto inspecting the sources of genetic gain by partitioning the genetic gain by paths. We also restructures and rewritten parts of the “Results” to better explain the findings. For each scenario we split the results into 1) comparing the genetic gain in nucleus vs multiplier 2) comparing the sources of genetic gain of the nucleus and multiplier.

------------------------------------  
L37-39 also a little hard to fully understand here.

We clarified the abstract by saying (lines 34-39): “*We partitioned the nucleus and multiplier genetic gain by tier‑gender path variable. For traits measured and selected in the multiplier, the multiplier selection generated additional genetic gain compared to nucleus. The increase depended on the accuracy and intensity of selection in the multiplier and the extent of gene flow from the nucleus. For traits measured only in the nucleus, the multiplier selection did not generate additional genetic gain.”*

------------------------------------  
L42 suggest "The use of AlphaPart will help breeders to improve genetic gain through a better understanding of the key selection points that are currently driving gains in each trait".

Thank you, we changed the manuscript to include the suggested sentence. We however did not say “currently”, since this helps to track the sources of genetic gain in the previous generations.

------------------------------------

L49 use "mating" instead of "intermating"

Replaced.  
------------------------------------  
L104 -117 would this be better presented as a table, first column giving the function name, and second column describing what the function does in words?

Thank you for this comment. We discussed the idea, but decided, that since we only present four functions, a table is not necessary.

------------------------------------  
L118. Needs to be clear that this is an open nucleus approach. It is not necessarily standard that elite individuals from the multiplier tier can make a genetic contribution to the nucleus tier?

We did not simulate an open nucleus, but a closed one. We added sentences throughout the manuscript to clarify this. For example, lines 181-183: “*We used the AlphaSimR package [10] to simulate the stylized multi‑tier breeding programme for a single breed with closed nucleus and directional flow of animals from the nucleus into the multiplier (Figure 2).*“ We also added a scheme of the breeding programme in Figure 2.

------------------------------------  
L120 replace "on" with "to"

Replaced.  
------------------------------------  
L122 supply "a" large number of

Added.  
------------------------------------

L135 it seems from the simulation that you are assuming one selection round per year. In reality, sows farrow multiple times per year, and there are lots of overlaps in selection

Thank you for this comment. We agree with the reviewer. We now changed “*years*” to “*generations*” and replaced “*pig breeding*” with “s*tylized multi‑tier breeding*” throughout the manuscript.

------------------------------------  
L138-143. It needs to be made clearer (maybe via a table of a figure) which out of multiplier and nucleus selections are being made, for each sex. Given that the multiplier is contributing genetic trend, I have the feeling from reading so far, that it is an open nucleus, so that elite males from the multiplier can contribute to nucleus matings?

We did not simulate an open nucleus. The selection of multiplier animals contributed to the genetic gain in the multiplier only. We also added a scheme of the breeding programme in Figure 2 to clarify the selection decision and flow of animals.

------------------------------------  
L143-145. I feel that this information is coming too late, at least for my taste. Also, the naming convention is not intuitive, and surely there is a huge difference in selection intensity, and why is selection across tiers depending on merit, rather than the apparent fixed proportions?

The scenarios are named by the percentage of flow of nucleus males to multiplier (everything else is the same). The MaleFlow100 has a 100% flow of males from nucleus to multiplier, so all multiplier females are inseminated with nucleus males. The MaleFlow20 has 20% of multiplier offspring from nucleus males and 80% of offspring from multiplier males. There is a difference in selection intensity of multiplier male selection, but it does not affect the naming of the scenarios.

and why is selection across tiers depending on merit, rather than the apparent fixed proportions?The selection in each tier is fixed in line with the described proportions. Every generation we select the same number of parents in the nucleus and the multiplier (described in lines 193-200) and inseminate a fixed proportion of multiplier females with nucleus males.

------------------------------------  
L147 …we ran "a" pedigree based…

Added.

------------------------------------  
L157..did you check to see if sources were the same for EBVs as for TBVs? This is an interesting question in itself which would be worth commenting on.

We estimated the EBVs and selected parents on the EBVs – while the partitionresults were qualitatively the same, we decided to show TBV results as these were clearer. We noticed a better match between the EBV and TBV for higher heritability trait. We highlight this in discussion (lines 327-330):

“*Users should take into account, that the accuracy of partitioning depends on the accuracy of the estimated breeding values and their Mendelian sampling terms, which are driven by the biology of the trait, breeding programme structure, and choice of the prediction model.*”

------------------------------------  
L171 "expectedly showed" is not good English

Thank you for the comment. We changed the sentence to (lines 238-239): “*As expected, the nucleus genetic gain stemmed completely from selection of nucleus males and nucleus females.”*