Response to reviewers.

We would like to express our gratitude to all reviewers for their valuable feedback. First, we would like to clarify an important point that may address some of the requests made: the marker panel we have developed **is specifically designed for kinship reassignment and is not intended for association studies**. To address this, we have added a clarifying paragraph in the Materials and Methods section.

That said, we've answered below every reviewer's question, which can lead to repetition.

REVIEW 1

In the materials and methods section, describe how to design organized mating plans (), I mean, what was the standard of the number of this number of birds (Muscovy breed or sire line and Pekin breed for dam line) or hybrid mule duck? Was the effective size of the population considered?

Here we're talking about experimental populations, which are, as is often the case, much smaller in effective size than commercial populations. In previous generations, we generally had 35 sires and 80 dams. In France, commercial poultry lines are subject to minimum breeding stock requirements, set at 30 sires and 150 dams for widely distributed duck populations (as opposed to populations with limited distribution, such as local breeds). These numbers result in an effective population size of Ne=100, according to Wright's (1931) formula, which is based on the sex ratio. With effective sizes of 101 (for Pekin ducks) and 111 (for Muscovy ducks), obtained with the same formula, we were close to the effective size obtained for a commercial population using the minimum numbers required in France. We, therefore, considered that our populations had sufficient effective size to provide a panel that could then be used in other populations.

How many parental lines were genotyped by high-density arrays (600K array)? Was a special statistical test considered for the number of selected people (power statistical test)?

This information is somewhat confidential. The paper related to this HD chip (Teissier et al, 2019) reads: “Sequence data for SNP discovery was generated from French duck populations including common and Muscovy ducks: ***several*** commercial lines from Grimaud Frères Sélection and Orvia Gourmaud Sélection, experimental lines from INRA, a mallard breed used as game bird, and Rouen duck.”

The authors are not specific about the number of populations they used.

In the materials and method section, at the beginning of how to select 192 markers, the number of remaining markers from 600 K is presented. How did this number of SNP remain? What was the meaning or purpose of the respected authors in evaluating the power of 96 SNPs?

The authors' aim is to build an inexpensive panel that can be used for both Pekin and Muscovy ducks, not forgetting the hybrid between the two (mule duck). They started from the SNP libraries present on the 600K chip and looked for common markers. Before doing so, they had to ensure that the markers present on the 600K chip (AXIOM technology with 35 bp primers) could be used in KASPar technology (50 bp primers). The paragraph has been reworded to make it clearer. **L103-118**

My suggestion is to evaluate the SNPs identified on specific chromosomes by checking for QTLs Database in these regions. The interesting thing to note in Figure 1 is the distribution of completely different allele frequencies in the two breeds, which needs to be discussed.

It should be remembered here that the SNPs used in a reassignment panel are chosen mainly for high MAF. Their position in a QTL is in no way a goal (the aim is not to select a trait that could be influenced by this QTL). On the contrary, a high MAF is rather the mark of a neutral locus, outside any area of the genome carrying a gene of interest.

As for the MAF distribution shown in Figure 1, it should be remembered that we are considering only 96 SNPs. It would seem difficult to discuss statistic distributions with so few observations. What's more, these allele frequencies were obtained from a relatively modest sample size, and in each breed the bulk of observations came from a single population.

Figure 2 shows the distribution of selected SNPs on different chromosomes. Respected authors have mentioned macrochromosomes 1 to 8, but in the figure, the number of markers on chromosome 16 and 20 is also high, which needs to be discussed.

This phenomenon is likely attributable to the selection of markers from the 399 SNPs common to the *Anas platyrhynchos* and *Cairina moschata* libraries. Although macrochromosomes are generally more represented on the Axiom chip, within the 399 shared SNPs, microchromosomes 9, 13 and 16 are more prevalent than macrochromosome 7. An additional figure has therefore been added, showing the distribution of SNPs on the Axiom HD chip and in the list of 399 SNPs in common. **L263-267**

It is still very unclear to me how to select 96 SNPs? Please provide the Reference or References or even(previous similar studies used. If possible, compare two genotyping technologies (Axiom and KASPar).

The aim is to reassign the target experimental populations. We have a panel of 192 SNPs in KASPar technology and would like to retain 96 (the capacity of a KASPar plate). To do this, we have two populations: a subgroup of the reference set (72 individuals: 44 Pekin, 15 Muscovy and 13 mule ducks) and the parents of the individuals to be reassigned (127 Pekin and 133 Muscovy ducks). The first group was used to check consistency between AXIOM and KASPar genotypes. In addition, it was made up of “animal+sire +dam” triplets to check that the panel was enabled us to find the parents. At the very least, this configuration has enabled us to search for any Mendelian incompatibilities, leading to the elimination of a marker. The second group was used to select markers on MAFs estimated in a population close to the target population.

In the materials section, the method of collecting eggs and hatching has been done, but there is no mention of this in the results and discussion? Please, add the result, as well as analyze the Association between these traits and the selected SNPs.

We are sorry, but we may not fully understand the question. As we have now clarified in the article, our objective was to develop an assignment panel, not to conduct an association study.

In the table 1, the Minor allelic frequency and call rate is different in two breeds, so that in breed 1 it is twice as much as breed 2. Please explain the reason for it according to the pedigree

and the breeding program and the duration of the breeding herds? Also discuss the low minimum allelic frequency in two breeds?

Clarification is needed here. In Table 1 we do not present two breeds for comparison, but describe the elementary statistics of the 192 SNP panel (in terms of call rate and MAF) in the two populations where we wish to be able to use the panel. These two populations belong to different genera that diverged several million years ago. That said, it doesn't seem to us that the distributions are very different. For the MAF, and even more so for the Call-Rate, the medians are very close (0.338 vs. 323) and (0.993 vs. 992) for Pekin and Muscovy ducks, respectively. In fact, the minimum Call-Rate value is abnormally low for Muscovy ducks (0.258), and this is explained in the text of the article: 57 markers did not respond in the Muscovy population. We assume that this is due to a polymorphism in the primer (going from 35 to 50 bp) that remained undetected in our starting sample, where the line at the base of our experimental population was only sparsely represented. With 57 deficient markers, the minimum value for MAF in Muscovy ducks (0.047) cannot be validly compared with the value of 0.026 found in Pekins.

My suggestion is to calculate the amount of linkage disequilibrium chromosome by chromosome. The number of SNPs after quality control should be presented and discussed in a table, chromosome by chromosome.

In chips used for genetic analysis, marker resolution is important, as there needs to be sufficient LD between a marker on the chip and the mutation we're trying to calculate. Here, the chip developed is an assignment chip, with a very different design prerequisite. LD calculation is, therefore, not necessary.

REVIEW2

I would like to express my sincere appreciation for the opportunity to review the preprint entitled "Cost-efficient assignment panel for ducks: Setup of a cost-efficient assignment panel for duck populations" with the potential for publication.

I regret the delay in my response, as I was deeply engrossed in my other commitments. Below:

I have outlined a comprehensive and detailed summary of my feedback and recommendations for the manuscript:

1. The research scenario is captivating, particularly in light of the challenges posed by commercial SNP chip panels for diverse breeds. I highly recommend that the authors emphasize the specific benefits of the panel for Pekin and Muscovy ducks and provide further insights into why these two breeds were specifically chosen for the study. Exploring the historical background and genetic relatedness of these breeds could add depth to the research.

The introduction has been revised to include a paragraph **(L54-62)** on the economic importance of the two populations studied here, recalling their distinct geographical origins and the age of their phylogenetic divergence. This increases the difficulty of developing a common tool for these two populations.

2. The abstract is well-crafted and informative. However, I suggest providing more comprehensive details about 192 SNP in the initial process and elucidating the relationship between SNP 96, 134, and 128 SNP. Additionally, including the population sizes of Muscovy and Pekin ducks in the abstract would augment its informativeness and provide a clearer context for the study.

There's some confusion here. While 192 and 96 are indeed the SNP numbers used in the panels developed during this study, the numbers 134 and 128 (which have now been corrected to 133 and 127) are the number of parents used in the Peking and Muscovy populations respectively). These numbers are now in the abstract.

3. Accurately reflecting the diversity is crucial, and it is essential to specify the representation of the panel across the chromosomes of Pekin and Muscovy ducks. Providing detailed information on the distribution of SNP along all chromosomes of poultry genomes will be essential to reflect the true face of diversity.

The distribution of SNPs across the chromosomes is provided in Table 2 for the 96 selected markers (and in the supplementary material for the 192 markers). A new paragraph **(L248-252)** now mentions the observed average and median distances between SNPs in both marker sets. The markers are distributed across the entire genome. Ultimately, the ability of the panel to reconstruct the pedigree of our target experimental populations serves as evidence that the markers are sufficiently well distributed.

4. The introduction would benefit from expansion, providing more in-depth background information on Pekin and Muscovy ducks and delving into the motivations behind the research. Highlighting the historical and genetic significance of these duck breeds and the specific research problems they address would add substantial value to the introduction.

As mentioned earlier, the introduction has been rephrased, and these pieces of information were included. **L54-62**

5. The "Materials and Methods" section is comprehensive and well-articulated. However, I recommend providing an accessible and detailed explanation of the KASpar technology for the benefit of the readers. Explaining the technology and its relevance to the study in an understandable manner will enhance the comprehension of the methods employed.

The section dedicated to the selection of 192 SNP now starts with a paragraph presenting the two technologies, and an additional figure comparing the KASPar and Axiom has been added. **L103-107**

6. The discrepancy in minor allele frequency values in Table 1 prompts the question of whether the number of suggested SNPs in the panel could be reduced without compromising genotyping outcomes. The authors should provide a robust justification for this information, including a detailed discussion on the implications of altering the suggested SNP numbers and its potential impact on genotyping outcomes.

In our opinion, the MAF values between the two populations do not exhibit significant differences, except for the minimum MAF, which is 0.026 in Pekin ducks compared to 0.047 in Muscovy ducks. The

other quantiles are comparable. However, the minimum MAF of 0.047 in the Muscovy population should be interpreted with caution due to the unusually low Call Rate observed in this population.

7. The results section is articulate and insightful. I found the discussion in this section to be enlightening and thought-provoking, and I appreciate the authors' thoroughness in presenting the results.

8. The conclusion should be more nuanced, offering a specific and concise summary of the outcomes. Providing a detailed and comprehensive summary of the outcomes, along with their implications and potential future directions, will enhance the conclusion section.

The conclusion has been revised, with the addition of a paragraph highlighting the potential applications of the intermediate panel developed prior to selecting our final set of 96 SNPs. **L452-455**

9. The references are meticulously cited and encompassing, and I appreciate the authors' attention to detail and the inclusion of a comprehensive range of updated references. In conclusion, with some minor revisions, I firmly believe that the manuscript is poised for publication.

I appreciate your attention to these detailed and comprehensive suggestions and hope that they contribute to the further enhancement of the manuscript.

With the Best Regards

Arash Javanmard

REVIEW 3

While I found the manuscript a little bit confusing at time (I would have described the factorial design before the SNP selection, lots of different numbers), the results of this research could be very useful in the future.

Could you add some information about KASpar here. If not, I am not sure why you are privileging this technology over Axiom, which has been used for the 600K? Why makes it low cost? What are the advantages?

Details regarding these two technologies have been added to the text and included in a figure in the supplementary material. **L103-107**

L73: What is your reasoning behind using 96 SNPs (why not 100?)

Our goal was to develop a cost-effective panel. Our partner laboratory proposed the KASPar technology, with plates capable of processing 96 SNPs. Consequently, the panel needed to contain a multiple of 96 SNPs. Naturally, we anticipated higher assignment rates with 192 SNPs, and the purpose of the experiment was to determine whether 96 SNPs would be sufficient. This was indeed the case

L78: I don't understand where the "only15% originated from same populations a parental lines". In line 72-73, you described the experimental lines as Cairina moshata and Anas platyrhynchus which are the same as the ones from the 600K.

The original quote was "15% originated from the same populations as OUR parental lines". We know which populations (commercial, experimental and heritage) were used to build the 600K chip. We also know the origin of our experimental lines. Given the great diversity between populations, even within the same breed (*i.e.* within Pekins ducks and within Muscovy ducks), we needed to estimate the allele frequencies in our populations to obtain the optimum panel of 96 SNPs. This led to a 2-stage strategy, which proved to be a winning one, given the difficulties encountered with some markers in our Muscovy population.

L81: What makes a SNP eligible for the "chosen technology" (I guess Kaspar)?

The sequence of steps leading to the 192 SNPs, as well as the conditions to be met for them to be "eligible", is hopefully now better described in **L103-127**. For clarification, "Chosen technology" has been replaced by KASPar

L82: "the first set" you have not describe what you mean by set prior to this sentence.

Our quote (L80) was: "In a first instance, a set of 192 SNPs eligible for the chosen technology was developed, based on both their frequencies in the three populations and their technical properties. This first set …".

The "first set" refers to the 192 SNPs, which are only an intermediate step in obtaining the final set of 96 markers.

L97: If only the SNPs with identical primers in both populations were kept, the number of remaining SNPs in each populations shouldn't be the same? (i.e. the intersection of the SNPs list between the 2 pop?). Confusingly, you are talking about the intersection of the 2 SNP list in the next sentence with yet another number 399.

Thank you, you are correct. The paragraph, as originally phrased, was indeed confusing. We have revised it to eliminate any ambiguity**. L108-118**

L115: where do these parents came from? Have they been genotyped prior the experiment? With Axiom? They cannot be from the reference dataset as you mention only 79 Muscovy for that one?

The parents referenced in line 115 are experimental animals that were not available at the time the 600K chip was constructed. The sole genotypes of these animals were obtained through the use of the 192 markers of the KASPar technology.

The objective was to construct the pedigree of three experimental herds, comprising Muscovy, 273 Pekin, and 207 mule ducks. To achieve this, we initially employed a reference set comprising 600K genotypes of *Anas platyrhynchos* (n=139), *Cairina moschata* (n=79), and some mule ducks (n=45). However, as these populations were not inherently related to our experimental subjects, we utilized the genotypes of the parents of the target batches to construct the final list of 96 markers, which were selected based on their suitability for use in our experimental design.

L154: In the factorial mating that you are proposing, the males were kept in individual cages. Will this design possible if the ban on cages in enforced?

The ban on cages would be motivated by respect for animal welfare. We believe that legislators will take account of aggressive behavior that is incompatible with the cohabitation of males in an enclosure. In this design, the females are on the ground in groups. All that remains is to house the males in compartments that comply with UE 2010/63 regulations.

Moreover, it seems that the European Commission has made no progress on this issue, which appeared to be a burning one, with legislation expected in 2027.

L160: Please explain your reasoning here

It's a matter of counting the possible cases. If we don't know the mating plan, a randomly drawn egg could come from 4608 Muscovy pairs (48 males crossed by 96 females). If we know the mating plan, this number is reduced to 384. A similar phenomenon is observed in Pekin

L162: Do you mean that you are avoiding to put sibs in the same group as their genotype may be too similar and the parentage assignment software may not be able to distinguish between them? If so, rephrase for better understanding

The paragraph was rephrased. **L181-183**

L184: I am a bit confused with the number here. In you mating design you use 96 +48 (144) Muscovy and 99 + 40 Pekin (139). Why do you genotype 157 Muscovy and 273 Pekin? The number of mules is only 207? Where these individuals already genotyped? With what technology?

These figures (157 Muscovy, 273 Pekin and 207 mule ducks) concern animals to be reassigned with the KASPar 96 SNP panel. Unlike the parents (133 Pekin ducks and 127 Muscovy ducks, these numbers have been adjusted in the corresponding paragraph), which were genotyped with all 192 SNP markers (before selecting the 96 most suited to our populations).

L209: Table 1. Any explanation as to why the minimum call-rate for Muscovy is so low (0.258) while minimum call rate for Pekin is 0.94?

As stated L217, “Fifty-seven SNPs exhibited missingness rate ranging from 0.42 to 0.75, while they were below 5% in the Muscovy samples previously genotyped with the 600K chip”. A missingness of 0.75 corresponds to a minimum call-rate of 0.24, rounded to the nearest two-digit number.

L264: What happened to theses 16 missing parental genotype? Where they genotyped but discarded because of poor quality or were they never genotyped? If not genotyped, why?

It is unfortunately a mixture of both. There was some confusion between the laboratory and the experimental unit, located 300 km apart. Consequently, the samples of these animals were not sent to the laboratory for genotyping. Once the anomaly was detected, indicated by abnormally low assignment rates, we made a one-off attempt to genotype the missing parents with the available funds. Unfortunately, the resulting genotypes were of poor quality, characterized by low Call-Rates which could be due to poor DNA extraction, and we lacked the resources to reattempt genotyping.