THE ROYAL SOCIETY PUBLISHING

PROCEEDINGS B

Highly heritable and functionally relevant breed differences in dog behaviour

Evan L. MacLean, Noah Snyder-Mackler, Bridgett M. vonHoldt and James A. Serpell

Article citation details
Proc. R. Soc. B **286**: 20190716.
http://dx.doi.org/10.1098/rspb.2019.0716

Review timeline
Original submission: 26 March 2019
1st revised submission: 29 May 2019
2nd revised submission: 6 September 2019
Final acceptance: 9 September 2019

Note: Reports are unedited and appear as submitted by the referee. The review history appears in chronological order.

Review History

RSPB-2019-0716.R0 (Original submission)

Review form: Reviewer 1

Recommendation
Major revision is needed (please make suggestions in comments)

Scientific importance: Is the manuscript an original and important contribution to its field?
Good

General interest: Is the paper of sufficient general interest?
Good

Quality of the paper: Is the overall quality of the paper suitable?
Good

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
No

Is it clear?
N/A

Is it adequate?
N/A

Do you have any ethical concerns with this paper?
No

Comments to the Author
I think that this manuscript has substantial potential to become a very nice contribution to Proc B. However, in my opinion, it would be much better if substantially fleshed out, possibly focusing on a subset of the major components that are currently presented. Following fleshing out, it seems to me that further review wold be necessary, as the amount of detail for a reviewer to go on is currently quite minimal in some aspects.

Major

(1) This paper is so brief that it is very difficult to know what context the authors intend for each component analysis - particularly, what patterns might be expected, and which, if any, might be surprising. To a similar extent, the manuscript is quite vague on what is to be made of different results. While I don't presume to know exactly the history of the paper, I think the authors would have to forgive the reader for wondering if this is a hasty re-submission following an initial try at a journal requiring a much shorter format. While I recognise (and like) that PSRB is flexible in the format of initial submissions, I feel that, with the many related results given in this manuscript, the authors have failed to make good use of the space afforded by PSRB lay out the background and interpretation of the various components of their paper. I think that the "among-breed heritability" and the GWAS components could be more fully described, for sure, but the additional studies beyond these two most developed components have such minimal treatment that the really feel like add-ons. I suspect that a fully-developed treatment of just the heritability-like analyses plus the GWAS could be a the basis fo a nice paper.

(2) I suggest that the use of the word "heritability" should be avoided, and also, the relationship between what is currently referred to as "heritability" at the among-breed level to heritability within breeds should be carefully articulated. Across biology, "heritability" is used here and there to denote various aspects of the degree of genetic control of phenotype, generating lots of potential confusion. In evolutionary quantitative genetics, "heritability" refers to the proportion of phenotypic variation that is attributable to (additive) genetic effects. This proportion is what is calculated here at the among breed level, but it is the level that I think requires reconsideration.

What the proportion heritability determines is the proportional response to selection (i.e., how much of the selection differential is represented by the response to selection) when selection is applied to some population to which selection might plausibly be applied. It is a property of some (essentially random mating) population at a given time, in a given environment. As such, heritability — in the context of the evolutionary QG principles that gives heritability its meaning — is the within-breed proportion of genetic variance. An analogous proportion calculated at some other level of biological organisation is fundamentally a different quantity. The calculation that has been performed is closely analogous (a proportion of variance), but I think that the biological meaning is so different that (a) a different word ought to be used, and (b) there paper needs in some way to reflect this difference.

This issue of wording and the theory underlying the notion of heritability comes back to my point (1) as well — what did the authors expect, and why, for the magnitude of the proportion of among-breed phenotypic differences that are genetically based? What did they expect for the difference between the existing within-breed heritability estimates and the new across-breed estimates of the proportion of genetic effects, and how is the difference to be interpreted? Are they to be thought of as the same thing (I think not, but to some extent the authors seem to), but the current one is somehow a better, bigger number? If they are not to be understood as representing the same thing, then what is to be understood from their comparison?

(3) Having laid out my points (1) and (2), I can try to articulate my point (3), which is really a hybrid of the first two points. If more space is taken to outline the background and some sorts of expectations, what would those expectations look like? On the main issue the magnitude of the heritability-like proportion calculated at the among breed level, are the qualitative findings expected or unexpected? I think they are exactly as we would expect. Regardless of the exact level of heritability, in the strict sense, so long as h^2 isn't trivial, then sustained diversifying selection, as has inarguably happened in dogs, will eventually generate large genetically-based differences among breeds. Taken long enough (and surely in dogs it has progressed long enough, considering the various reasons why it is obvious anyway that there are massive genetic differences, e.g., two breeds both raised in the common environment my kitchen look and in many ways act totally different from each other but perfectly typical for their breed, hybrids are often very much intermediate in many characters including temperament, etc…), then the proportion of among breed variation that is genetic will inevitably exceed the proportion of within-breed variation that is genetic. Similarly, if there are genetic differences that are picked out by the linear modelling analysis, then, power issues notwithstanding, one would hope that at least some of the underlying loci, which must necessarily exist, could be detected (and this seems to be the case). I do not necessarily think that the fact that everything might be neatly as expected is a bad thing — so much evolutionary QG gets massively and inconveniently complicated, that a neat study laying out what we should expect in a system like dogs, and showing that things do indeed fit together, could be a lovely thing.

(4) The fleshing out that I'm suggesting may not necessarily be entirely presentational. I've been trying to think a bit about what more carefully drawn links to underlying theory and expectations might look like. While the current discussion give little explicit guidance on what to think about the high degree of genetic determination of breed differences, it seems that the intention is that the reader sees this as a signature of responses to diversifying selection. However, drift alone would generate variation among breed means in proportion to the degree of genetic differentiation among breeds. This is exactly the signal that is quantified statistically in the GRM-mixed model analyses. The effects of selection are actually represented in whatever genetic differences exist, over and above those that would arise under drift. I suspect, given how recently the breeds have diverged, and despite the likely bottlenecks in the development of many breeds, that the diversification far exceeds what would happen under drift - but some kind of F_ST vs Q_ST approach, or better yet, calculations (based on admittedly a certain amount of

guesswork e.g., number of generations, effective population sizes, etc.; even quite rough back-of-the-envelop work could be really interesting), would be necessary to relate the mixed model results to the degree to which the genetic basis of among-breed differences can be related to different evolutionary processes.

(5) While the methods do cover all analyses, I think that they could stand to be fleshed out quite a lot too. Many analyses are described via the statistical packages that were employed, and in some cases the names of the options used in those packages. In five or ten years, many of these packages will be obsolete and people will have forgotten (and in many cases will never have known in the first place) what these packages and options did. I think that descriptions need to be given much more of what the actual biostatistics were. For e.g., (this is not exhaustive) (i) what measure of genomic relatedness was used to make the genomic relatedness matrix (there are multiple potential options, and even if your choice was quite standard, what is standard now may not be in five years time)? (ii) What were the breeds of the four dogs in the gene expression study (and does it matter, or more importantly, as it doesn't seem to matter massively to the authors, why doesn't it?)? (iii) What is the reader supposed to understand from the one bit of detail about "raw questionnaire item scores, rather than factor scores" (line 264) - this seems to assume knowledge about the data that the reader does not have?, (iv) of the 25 individuals per breed that were subsampled for each mixed model run, were their phenotypes averaged, or used as separate measures? If the former, say so, if the latter, then there must be an additional variance component in the model (giving an equation for the mixed model would be a great way to help flesh out the description and solve this kind of ambiguity).

(6) There seems - but it certainly could be my misunderstanding - to be an unnecessary convolution in the estimation of the heritability-like parameter. I gather from 246-249 that many analyses were done, using just a few records per breed. The reason for this is not given, but from first principles of mixed modelling, this should not be necessary. A mixed model should be able to use the raw individual-level data, and partition that into a residual variance (differences among individuals, within breeds), and breed-level variance components that are and are not associated with the genetic differences among breeds; the heritability-like ratio constructed from these latter two components would be the proportion of breed differences that is due to genetic differences among breeds. This would be a lot neater, and would make full use of all the data in a single analysis. If this analysis was too slow because of the many individual records, then a mean and SE for each breed could be calculated, and these could be used as meta-analytic data in MCMCglmm (and some REML softwares could also accommodate this alternative parameterisation), wherein there could be a GRM-linked variance component, and the residual variance would represent among-breed, non-GRM related, differences among breeds.

(Signed; n.b. all my reviews are signed) Michael Morrissey

Specific

- line 54: two points. I would consider dropping the "compared to humans" - at least for PRSB, I would think that the biological principles, irrespective of species, are of interest. I would delete the word "uniquely" (also l.203-204) - I don't think it is true, an whether it is true is probably not relevant.

- line 58: how a "limited genetic toolkit"? What is a "genetic toolkit"? I feel this kind of phrasing is common, but just muddies things, as there is really no specific meaning. I'd remove.

- l.82 "control for differences among breeds" this is the kind of phrasing that works best if we are trying to control for some confounding variable. Actually, the relatedness among breeds is they key source of information here, not some nuisance!

- l.87: in my opinion, it would be much better to give means and standard errors (or better yet, differences, and SEs of the differences) rather than test statistics and p values without any notion of biological effects, or uncertainty in those effects. This applies throughout.

- l.93: "cultivation" seems a strange word

- l.94/95. Yes, there is undoubtedly a correspondence between the genetic and phenotypic clustering. But there is no sense in which they recapitulate one another with the "high accuracy" that you describe. The ability to predict one from the other is very vague. Also, this might be one of the best of many opportunities to reinforce my main point (1). The many potentially interesting facets of the work that went into figure 1C are represented by only this one statement (which I believe to be superficial and over-stated). I think there is a potentially a lot of value here that is greatly suffering from the extremely abbreviated presentation.

I.119: Bonferroni doesn't (classically) change the p value, but rather the critical p value (a.k.a. the alpha value). It seem unnecessarily confusing to state it this way, as though the p values themselves have somehow been corrected.

I.26-128. Calculating the variance associated with top SNPs is actually a really difficult thing, because, on average, the *estimated* effect sizes of the top SNPs will surely be (on average) overestimates of the *true* effects of the top SNPs. This is the GWAS analogue of the Beavis effect in QTL analysis. I'm not sure if your calculations here try to control in some way for this effect (i.e., if it is anything more than a sum over loci of 2pq\alpha), or even what your calculations are. It is another case of much more methodological detail being necessary before a real evaluation can be made (ether by a reviewer, or more importantly, a reader).

l.130-134, l.147/8. The supplemental stuff referred to here seems to me to actually be the important "meat" of a major component of this manuscript. I think it could be much more developed - maybe even into a manuscript of its own! (n.b., I hope you won't think I'm advocating for minimal publishable units here — I think there really could be multiple meaty papers here!)

From l.169: way to little detail here, in my opinion.

l.194: I would not describe this as a "limitation". The information needed to study differences among breeds, which is something entirely worth studying, is well-provided by the data to hand!

Review form: Reviewer 2 (Rosalind Arden)

Recommendation
Accept with minor revision (please list in comments)

Scientific importance: Is the manuscript an original and important contribution to its field?
Excellent

General interest: Is the paper of sufficient general interest?
Excellent

Quality of the paper: Is the overall quality of the paper suitable?
Excellent

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
Yes

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
Yes

Is it clear?
Yes

Is it adequate?
Yes

Do you have any ethical concerns with this paper?
No

Comments to the Author
To preserve paragraph spaces (which were eliminated in the preview of my comments) I have attached a WORD doc (See Appendix A).

Decision letter (RSPB-2019-0716.R0)

29-Apr-2019

Dear Dr MacLean:

Your manuscript has now been peer reviewed and the reviews have been assessed by an Associate Editor. The reviewers' comments (not including confidential comments to the Editor) and the comments from the Associate Editor are included at the end of this email for your reference. As you will see, the reviewers and the Editors have raised some concerns with your manuscript and we would like to invite you to revise your manuscript to address them.

We do not allow multiple rounds of revision so we urge you to make every effort to fully address all of the comments at this stage. If deemed necessary by the Associate Editor, your manuscript will be sent back to one or more of the original reviewers for assessment. If the original reviewers are not available we may invite new reviewers. Please note that we cannot guarantee eventual acceptance of your manuscript at this stage.

To submit your revision please log into http://mc.manuscriptcentral.com/prsb and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with

Decisions." Under "Actions", click on "Create a Revision". Your manuscript number has been appended to denote a revision.

When submitting your revision please upload a file under "Response to Referees" - in the "File Upload" section. This should document, point by point, how you have responded to the reviewers' and Editors' comments, and the adjustments you have made to the manuscript. We require a copy of the manuscript with revisions made since the previous version marked as 'tracked changes' to be included in the 'response to referees' document.

Your main manuscript should be submitted as a text file (doc, txt, rtf or tex), not a PDF. Your figures should be submitted as separate files and not included within the main manuscript file.

When revising your manuscript you should also ensure that it adheres to our editorial policies (https://royalsociety.org/journals/ethics-policies/). You should pay particular attention to the following:

Research ethics:
If your study contains research on humans please ensure that you detail in the methods section whether you obtained ethical approval from your local research ethics committee and gained informed consent to participate from each of the participants.

Use of animals and field studies:
If your study uses animals please include details in the methods section of any approval and licences given to carry out the study and include full details of how animal welfare standards were ensured. Field studies should be conducted in accordance with local legislation; please include details of the appropriate permission and licences that you obtained to carry out the field work.

Data accessibility and data citation:
It is a condition of publication that you make available the data and research materials supporting the results in the article. Datasets should be deposited in an appropriate publicly available repository and details of the associated accession number, link or DOI to the datasets must be included in the Data Accessibility section of the article (https://royalsociety.org/journals/ethics-policies/data-sharing-mining/). Reference(s) to datasets should also be included in the reference list of the article with DOIs (where available).

In order to ensure effective and robust dissemination and appropriate credit to authors the dataset(s) used should also be fully cited and listed in the references.

If you wish to submit your data to Dryad (http://datadryad.org/) and have not already done so you can submit your data via this link
http://datadryad.org/submit?journalID=RSPB&manu=(Document not available), which will take you to your unique entry in the Dryad repository.

If you have already submitted your data to dryad you can make any necessary revisions to your dataset by following the above link.

For more information please see our open data policy http://royalsocietypublishing.org/data-sharing.

Electronic supplementary material:
All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online

figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI. Please try to submit all supplementary material as a single file.

Online supplementary material will also carry the title and description provided during submission, so please ensure these are accurate and informative. Note that the Royal Society will not edit or typeset supplementary material and it will be hosted as provided. Please ensure that the supplementary material includes the paper details (authors, title, journal name, article DOI). Your article DOI will be 10.1098/rspb.[paper ID in form xxxx.xxxx e.g. 10.1098/rspb.2016.0049].

Please submit a copy of your revised paper within three weeks. If we do not hear from you within this time your manuscript will be rejected. If you are unable to meet this deadline please let us know as soon as possible, as we may be able to grant a short extension.

Thank you for submitting your manuscript to Proceedings B; we look forward to receiving your revision. If you have any questions at all, please do not hesitate to get in touch.

Best wishes,
Proceedings B
mailto: proceedingsb@royalsociety.org

Associate Editor
Comments to Author:
Thank you for submitting your work to PRSB. I have now received comments on your MS from two experts in the field and have read your paper carefully myself. We all agree that your study tackles an interesting question and that your paper shows a lot of potential. However, both reviewers converged on a concern that the paper is not as thorough or as clear as it it could be and both reviewers pointed to several areas that would benefit from more work (e.g., fleshing out the description of the design, adding more information about the methods, being careful and specific with terminology). To draw your attention to a few specific comments: Reviewer 2 asks that you provide more information about the behavioral traits so that readers can understand the basics of this method without having to read other papers while Reviewer 1 points to several areas that need substantial reworking, including laying out the background and expectations more clearly and potentially rethinking some of the analyses. Addressing these comments will require substantial revisions. However, both reviewers believe, as do I, that the paper will be stronger if you are able to address them. Please note that R1 has chosen to sign their review.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)
I think that this manuscript has substantial potential to become a very nice contribution to Proc B. However, in my opinion, it would be much better if substantially fleshed out, possibly focusing on a subset of the major components that are currently presented. Following fleshing out, it seems to me that further review wold be necessary, as the amount of detail for a reviewer to go on is currently quite minimal in some aspects.

Major

(1) This paper is so brief that it is very difficult to know what context the authors intend for each

component analysis - particularly, what patterns might be expected, and which, if any, might be surprising. To a similar extent, the manuscript is quite vague on what is to be made of different results. While I don't presume to know exactly the history of the paper, I think the authors would have to forgive the reader for wondering if this is a hasty re-submission following an initial try at a journal requiring a much shorter format. While I recognise (and like) that PSRB is flexible in the format of initial submissions, I feel that, with the many related results given in this manuscript, the authors have failed to make good use of the space afforded by PSRB lay out the background and interpretation of the various components of their paper. I think that the "among-breed heritability" and the GWAS components could be more fully described, for sure, but the additional studies beyond these two most developed components have such minimal treatment that the really feel like add-ons. I suspect that a fully-developed treatment of just the heritability-like analyses plus the GWAS could be a the basis fo a nice paper.

(2) I suggest that the use of the word "heritability" should be avoided, and also, the relationship between what is currently referred to as "heritability" at the among-breed level to heritability within breeds should be carefully articulated. Across biology, "heritability" is used here and there to denote various aspects of the degree of genetic control of phenotype, generating lots of potential confusion. In evolutionary quantitative genetics, "heritability" refers to the proportion of phenotypic variation that is attributable to (additive) genetic effects. This proportion is what is calculated here at the among breed level, but it is the level that I think requires reconsideration. What the proportion heritability determines is the proportional response to selection (i.e., how much of the selection differential is represented by the response to selection) when selection is applied to some population to which selection might plausibly be applied. It is a property of some (essentially random mating) population at a given time, in a given environment. As such, heritability — in the context of the evolutionary QG principles that gives heritability its meaning — is the within-breed proportion of genetic variance. An analogous proportion calculated at some other level of biological organisation is fundamentally a different quantity. The calculation that has been performed is closely analogous (a proportion of variance), but I think that the biological meaning is so different that (a) a different word ought to be used, and (b) there paper needs in some way to reflect this difference.

This issue of wording and the theory underlying the notion of heritability comes back to my point (1) as well — what did the authors expect, and why, for the magnitude of the proportion of among-breed phenotypic differences that are genetically based? What did they expect for the difference between the existing within-breed heritability estimates and the new across-breed estimates of the proportion of genetic effects, and how is the difference to be interpreted? Are they to be thought of as the same thing (I think not, but to some extent the authors seem to), but the current one is somehow a better, bigger number? If they are not to be understood as representing the same thing, then what is to be understood from their comparison?

(3) Having laid out my points (1) and (2), I can try to articulate my point (3), which is really a hybrid of the first two points. If more space is taken to outline the background and some sorts of expectations, what would those expectations look like? On the main issue the magnitude of the heritability-like proportion calculated at the among breed level, are the qualitative findings expected or unexpected? I think they are exactly as we would expect. Regardless of the exact level of heritability, in the strict sense, so long as h^2 isn't trivial, then sustained diversifying selection, as has inarguably happened in dogs, will eventually generate large genetically-based differences among breeds. Taken long enough (and surely in dogs it has progressed long enough, considering the various reasons why it is obvious anyway that there are massive genetic differences, e.g., two breeds both raised in the common environment my kitchen look and in many ways act totally different from each other but perfectly typical for their breed, hybrids are often very much intermediate in many characters including temperament, etc…), then the proportion of among breed variation that is genetic will inevitably exceed the proportion of

within-breed variation that is genetic. Similarly, if there are genetic differences that are picked out by the linear modelling analysis, then, power issues notwithstanding, one would hope that at least some of the underlying loci, which must necessarily exist, could be detected (and this seems to be the case). I do not necessarily think that the fact that everything might be neatly as expected is a bad thing — so much evolutionary QG gets massively and inconveniently complicated, that a neat study laying out what we should expect in a system like dogs, and showing that things do indeed fit together, could be a lovely thing.

(4) The fleshing out that I'm suggesting may not necessarily be entirely presentational. I've been trying to think a bit about what more carefully drawn links to underlying theory and expectations might look like. While the current discussion give little explicit guidance on what to think about the high degree of genetic determination of breed differences, it seems that the intention is that the reader sees this as a signature of responses to diversifying selection. However, drift alone would generate variation among breed means in proportion to the degree of genetic differentiation among breeds. This is exactly the signal that is quantified statistically in the GRM-mixed model analyses. The effects of selection are actually represented in whatever genetic differences exist, over and above those that would arise under drift. I suspect, given how recently the breeds have diverged, and despite the likely bottlenecks in the development of many breeds, that the diversification far exceeds what would happen under drift - but some kind of F_ST vs Q_ST approach, or better yet, calculations (based on admittedly a certain amount of guesswork e.g., number of generations, effective population sizes, etc.; even quite rough back-of-the-envelop work could be really interesting), would be necessary to relate the mixed model results to the degree to which the genetic basis of among-breed differences can be related to different evolutionary processes.

(5) While the methods do cover all analyses, I think that they could stand to be fleshed out quite a lot too. Many analyses are described via the statistical packages that were employed, and in some cases the names of the options used in those packages. In five or ten years, many of these packages will be obsolete and people will have forgotten (and in many cases will never have known in the first place) what these packages and options did. I think that descriptions need to be given much more of what the actual biostatistics were. For e.g., (this is not exhaustive) (i) what measure of genomic relatedness was used to make the genomic relatedness matrix (there are multiple potential options, and even if your choice was quite standard, what is standard now may not be in five years time)? (ii) What were the breeds of the four dogs in the gene expression study (and does it matter, or more importantly, as it doesn't seem to matter massively to the authors, why doesn't it?)? (iii) What is the reader supposed to understand from the one bit of detail about "raw questionnaire item scores, rather than factor scores" (line 264) - this seems to assume knowledge about the data that the reader does not have?, (iv) of the 25 individuals per breed that were subsampled for each mixed model run, were their phenotypes averaged, or used as separate measures? If the former, say so, if the latter, then there must be an additional variance component in the model (giving an equation for the mixed model would be a great way to help flesh out the description and solve this kind of ambiguity).

(6) There seems - but it certainly could be my misunderstanding - to be an unnecessary convolution in the estimation of the heritability-like parameter. I gather from 246-249 that many analyses were done, using just a few records per breed. The reason for this is not given, but from first principles of mixed modelling, this should not be necessary. A mixed model should be able to use the raw individual-level data, and partition that into a residual variance (differences among individuals, within breeds), and breed-level variance components that are and are not associated with the genetic differences among breeds; the heritability-like ratio constructed from these latter two components would be the proportion of breed differences that is due to genetic differences among breeds. This would be a lot neater, and would make full use of all the data in a single analysis. If this analysis was too slow because of the many individual records, then a

mean and SE for each breed could be calculated, and these could be used as meta-analytic data in MCMCglmm (and some REML softwares could also accommodate this alternative parameterisation), wherein there could be a GRM-linked variance component, and the residual variance would represent among-breed, non-GRM related, differences among breeds.

(Signed; n.b. all my reviews are signed) Michael Morrissey

Specific

- line 54: two points. I would consider dropping the "compared to humans" - at least for PRSB, I would think that the biological principles, irrespective of species, are of interest. I would delete the word "uniquely" (also l.203-204) - I don't think it is true, an whether it is true is probably not relevant.

- line 58: how a "limited genetic toolkit"? What is a "genetic toolkit"? I feel this kind of phrasing is common, but just muddies things, as there is really no specific meaning. I'd remove.

- l.82 "control for differences among breeds" this is the kind of phrasing that works best if we are trying to control for some confounding variable. Actually, the relatedness among breeds is they key source of information here, not some nuisance!

- l.87: in my opinion, it would be much better to give means and standard errors (or better yet, differences, and SEs of the differences) rather than test statistics and p values without any notion of biological effects, or uncertainty in those effects. This applies throughout.

- l.93: "cultivation" seems a strange word

- l.94/95. Yes, there is undoubtedly a correspondence between the genetic and phenotypic clustering. But there is no sense in which they recapitulate one another with the "high accuracy" that you describe. The ability to predict one from the other is very vague. Also, this might be one of the best of many opportunities to reinforce my main point (1). The many potentially interesting facets of the work that went into figure 1C are represented by only this one statement (which I believe to be superficial and over-stated). I think there is a potentially a lot of value here that is greatly suffering from the extremely abbreviated presentation.

I.119: Bonferroni doesn't (classically) change the p value, but rather the critical p value (a.k.a. the alpha value). It seem unnecessarily confusing to state it this way, as though the p values themselves have somehow been corrected.

I.26-128. Calculating the variance associated with top SNPs is actually a really difficult thing, because, on average, the *estimated* effect sizes of the top SNPs will surely be (on average) overestimates of the *true* effects of the top SNPs. This is the GWAS analogue of the Beavis effect in QTL analysis. I'm not sure if your calculations here try to control in some way for this effect (i.e., if it is anything more than a sum over loci of 2pq\alpha), or even what your calculations are. It is another case of much more methodological detail being necessary before a real evaluation can be made (ether by a reviewer, or more importantly, a reader).

l.130-134, l.147/8. The supplemental stuff referred to here seems to me to actually be the important "meat" of a major component of this manuscript. I think it could be much more developed - maybe even into a manuscript of its own! (n.b., I hope you won't think I'm advocating for minimal publishable units here — I think there really could be multiple meaty papers here!)

From l.169: way to little detail here, in my opinion.

l.194: I would not describe this as a “limitation”. The information needed to study differences among breeds, which is something entirely worth studying, is well-provided by the data to hand!

Referee: 2

Comments to the Author(s)
To preserve paragraph spaces (which were eliminated in the preview of my comments) I have attached a WORD doc.

Author's Response to Decision Letter for (RSPB-2019-0716.R0)

See Appendix B.

RSPB-2019-0716.R1 (Revision)

Review form: Reviewer 1

Recommendation
Major revision is needed (please make suggestions in comments)

Scientific importance: Is the manuscript an original and important contribution to its field?
Acceptable

General interest: Is the paper of sufficient general interest?
Acceptable

Quality of the paper: Is the overall quality of the paper suitable?
Acceptable

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
N/A

Is it clear?
N/A

Is it adequate?
N/A

Do you have any ethical concerns with this paper?
No

Comments to the Author
.

Review form: Reviewer 2

Recommendation
Accept with minor revision (please list in comments)

Scientific importance: Is the manuscript an original and important contribution to its field?
Excellent

General interest: Is the paper of sufficient general interest?
Excellent

Quality of the paper: Is the overall quality of the paper suitable?
Good

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
Yes

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
Yes

Is it clear?
Yes

Is it adequate?
Yes

Do you have any ethical concerns with this paper?
No

Comments to the Author
Line 7--72
"For example, although breed differences in behavior are well documented (reviewed in 3), it remains unknown to what extent these differences are attributable to genetic relatedness among breeds. "
Line 73-75
"First, we hypothesized that if diversifying selection in dogs has led to genetically-based breed differences in behavior, then much of the behavioral diversity among breeds should be attributable to genetic similarity among breeds. "

In both the above sentences, the reader has to work hard. In standard QG reports, the focus is on finding the extent to which phenotypic similarity is predicted by genetic similarity. Seeking diversity to explain similarity is hard work to untangle - could this be worded more intelligibly?

Line 92
Later you say "Second, if breed differences in behavior are largely attributable to genetic differences among breeds, ..." this has the virtue of both nouns (breed differences, genetic differences) pointing in the same direction.

Lind 78-80
When considering variance across breeds, we can make use heritability while acknowledging some potentially important differences from studies of heritability in randomly interbreeding populations

Typo 'of' missing in action?

Review form: Reviewer 3

Recommendation
Accept with minor revision (please list in comments)

Scientific importance: Is the manuscript an original and important contribution to its field?
Good

General interest: Is the paper of sufficient general interest?
Excellent

Quality of the paper: Is the overall quality of the paper suitable?
Excellent

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
Yes

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
No

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
N/A

Is it clear?
N/A

Is it adequate?
N/A

Do you have any ethical concerns with this paper?
No

Comments to the Author
I think this is a very good paper -- but the Abstract somewhat undersells the findings. I would find stronger terms than "provide insight" and "suggest that dogs provide a powerful model." This makes it sound like the papers main findings are methodological, whereas there are real findings here that people care about: Namely that we now have estimates of among-breed heritabilities of certain personality types; they are quite high. Further that the specific genes responsible for typical behavioral differences between breeds of dogs have been identified.

Decision letter (RSPB-2019-0716.R1)

02-Sep-2019

Dear Dr MacLean

I am pleased to inform you that your manuscript RSPB-2019-0716.R1 entitled "Highly Heritable and Functionally Relevant Breed Differences in Dog Behavior" has been accepted for publication in Proceedings B.

The referees have recommended publication, but also suggest some minor revisions to your manuscript. In particular, there were concerns shared in the confidential comments that your use of the term "heritability" is somewhat unconventional and so it would be helpful if you justified your terminology in the introduction somewhere. Therefore, I invite you to respond to the referees' comments and revise your manuscript. Because the schedule for publication is very

tight, it is a condition of publication that you submit the revised version of your manuscript within 7 days. If you do not think you will be able to meet this date please let us know.

To revise your manuscript, log into https://mc.manuscriptcentral.com/prsb and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision. You will be unable to make your revisions on the originally submitted version of the manuscript. Instead, revise your manuscript and upload a new version through your Author Centre.

When submitting your revised manuscript, you will be able to respond to the comments made by the referee(s) and upload a file "Response to Referees". You can use this to document any changes you make to the original manuscript. We require a copy of the manuscript with revisions made since the previous version marked as 'tracked changes' to be included in the 'response to referees' document.

Before uploading your revised files please make sure that you have:

1) A text file of the manuscript (doc, txt, rtf or tex), including the references, tables (including captions) and figure captions. Please remove any tracked changes from the text before submission. PDF files are not an accepted format for the "Main Document".

2) A separate electronic file of each figure (tiff, EPS or print-quality PDF preferred). The format should be produced directly from original creation package, or original software format. PowerPoint files are not accepted.

3) Electronic supplementary material: this should be contained in a separate file and where possible, all ESM should be combined into a single file. All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI.

Online supplementary material will also carry the title and description provided during submission, so please ensure these are accurate and informative. Note that the Royal Society will not edit or typeset supplementary material and it will be hosted as provided. Please ensure that the supplementary material includes the paper details (authors, title, journal name, article DOI). Your article DOI will be 10.1098/rspb.[paper ID in form xxxx.xxxx e.g. 10.1098/rspb.2016.0049].

4) A media summary: a short non-technical summary (up to 100 words) of the key findings/importance of your manuscript.

5) Data accessibility section and data citation
It is a condition of publication that data supporting your paper are made available either in the electronic supplementary material or through an appropriate repository.

In order to ensure effective and robust dissemination and appropriate credit to authors the dataset(s) used should be fully cited. To ensure archived data are available to readers, authors should include a 'data accessibility' section immediately after the acknowledgements section. This should list the database and accession number for all data from the article that has been made publicly available, for instance:
• DNA sequences: Genbank accessions F234391-F234402
• Phylogenetic data: TreeBASE accession number S9123

• Final DNA sequence assembly uploaded as online supplemental material
• Climate data and MaxEnt input files: Dryad doi:10.5521/dryad.12311
NB. From April 1 2013, peer reviewed articles based on research funded wholly or partly by RCUK must include, if applicable, a statement on how the underlying research materials – such as data, samples or models – can be accessed. This statement should be included in the data accessibility section.

If you wish to submit your data to Dryad (http://datadryad.org/) and have not already done so you can submit your data via this link
http://datadryad.org/submit?journalID=RSPB&manu=(Document not available) which will take you to your unique entry in the Dryad repository. If you have already submitted your data to dryad you can make any necessary revisions to your dataset by following the above link. Please see https://royalsociety.org/journals/ethics-policies/data-sharing-mining/ for more details.

6) For more information on our Licence to Publish, Open Access, Cover images and Media summaries, please visit https://royalsociety.org/journals/authors/author-guidelines/.

Once again, thank you for submitting your manuscript to Proceedings B and I look forward to receiving your revision. If you have any questions at all, please do not hesitate to get in touch.

Sincerely,

Dr Sasha Dall
Editor, Proceedings B
mailto:proceedingsb@royalsociety.org

Reviewer(s)' Comments to Author:

Referee: 2

Comments to the Author(s)
Line 7--72
"For example, although breed differences in behavior are well documented (reviewed in 3), it remains unknown to what extent these differences are attributable to genetic relatedness among breeds. ?"
Line 73-75
"First, we hypothesized that if diversifying selection in dogs has led to genetically-based breed differences in behavior, then much of the behavioral diversity among breeds should be ? attributable to genetic similarity among breeds. ?"

In both the above sentences, the reader has to work hard. In standard QG reports, the focus is on finding the extent to which phenotypic similarity is predicted by genetic similarity. Seeking diversity to explain similarity is hard work to untangle - could this be worded more intelligibly?

Line 92
Later you say "Second, if breed differences in behavior are largely attributable to genetic differences ?among breeds, ?...." this has the virtue of both nouns (breed differences, genetic differences) pointing in the same direction.

Lind 78-80
When considering variance across breeds, we can make use heritability while acknowledging some potentially important differences from studies of heritability ? in randomly interbreeding populations ?

Typo 'of' missing in action?

Referee: 3

Comments to the Author(s)
I think this is a very good paper -- but the Abstract somewhat undersells the findings. I would find stronger terms than "provide insight" and "suggest that dogs provide a powerful model." This makes it sound like the papers main findings are methodological, whereas there are real findings here that people care about: Namely that we now have estimates of among-breed heritabilities of certain personality types; they are quite high. Further that the specific genes responsible for typical behavioral differences between breeds of dogs have been identified.

Author's Response to Decision Letter for (RSPB-2019-0716.R1)

See Appendix C.

Decision letter (RSPB-2019-0716.R2)

09-Sep-2019

Dear Dr MacLean

I am pleased to inform you that your manuscript entitled "Highly Heritable and Functionally Relevant Breed Differences in Dog Behavior" has been accepted for publication in Proceedings B.

You can expect to receive a proof of your article from our Production office in due course, please check your spam filter if you do not receive it. PLEASE NOTE: you will be given the exact page length of your paper which may be different from the estimation from Editorial and you may be asked to reduce your paper if it goes over the 10 page limit.

If you are likely to be away from e-mail contact please let us know. Due to rapid publication and an extremely tight schedule, if comments are not received, we may publish the paper as it stands.

If you have any queries regarding the production of your final article or the publication date please contact procb_proofs@royalsociety.org

Your article has been estimated as being 10 pages long. Our Production Office will be able to confirm the exact length at proof stage.

Open Access
You are invited to opt for Open Access, making your freely available to all as soon as it is ready for publication under a CCBY licence. Our article processing charge for Open Access is £1700.

Corresponding authors from member institutions
(http://royalsocietypublishing.org/site/librarians/allmembers.xhtml) receive a 25% discount to these charges. For more information please visit http://royalsocietypublishing.org/open-access.

Paper charges
An e-mail request for payment of any related charges will be sent out shortly. The preferred payment method is by credit card; however, other payment options are available.

Electronic supplementary material:
All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI.

Thank you for your fine contribution. On behalf of the Editors of the Proceedings B, we look forward to your continued contributions to the Journal.

Sincerely,

Editor, Proceedings B
mailto: proceedingsb@royalsociety.org

Appendix A

Review of Maclean et al

Two questions are posed and answered in this ms: 1) to what extent are behavioural traits heritable in dogs? 2) If heritable, what is the genetic architecture of the traits? Are they largely polygenic?

This paper is a corker. The questions are worth asking, the research design provides a clever way to probe the questions, absent individual level behavioural and genetic data on the same dogs. The interpretation of the results sticks well to the data without over-reaching. It's an ambitious piece of work that looks well executed and mostly very well written. It will make an excellent contribution to the literature.

I have some comments for the Authors to consider:

Spell out the essential design of a little more clearly. I thought that since this paper does not use the standard behavioural genetic designs to explore answers to the key question about trait heritability, it would be helpful to further elaborate the design. The design is so thoroughly well-known to the Authors; it is hard to escape the curse of knowledge. For example, readers may wrongly think that there is a breed specific reference genome for each of the 101 breeds in the study. It wouldn't take much work by the Authors to allow readers to follow a little more easily what was done.

The Results begin with Methods, I'd consider starting with " We found that... and moving the part before it.

The 14 behavioural traits.
What I wanted to know for each of these was: an example showing how the trait scores are derived and an index of the reliability of each trait. The heritability of mush is less interesting than the heritability of a well-measured trait so it's important to know something more about the measures. In line 90 you write about h2 estimates - is h2 correlated with reliability? Are more reliably measured traits more heritable? Many readers will not know C-BARQ - I assume some traits are more reliably assessed than others. Could this be given in the supplementary materials so that readers do not have to first research C-BARQ?

Fig1 B and C
I printed these out and also looked at them on screen so I could increase the size of the text. I couldn't read them either way. Consider a note to say that a larger version is provided in Supplementary materials?

Fig2 A
The colour index is a little too fine grained- one suggestion is that a Manhattan plot for selected traits is presented in main text with a note that Manhattan plots for other traits are in Supplementary. Just an idea.

The trait 'Energy'. I don't know what this means. Please can it be explained?
Line 153. Assertion needs a citation (genes associated with intelligence and information processing).

Line 86-87: worth mentioning range-restriction - to explain the difference in estimates? I'd lose the "identified" it's redundant here.

Line 191: BG is a set of methods, so I'm unsure what is meant by variant being "implicated in human BG" do you mean that there may be the same variants may also be associated with tameness/ aggression in humans? If so, perhaps make clearer and give a citation.

194-197 Limitations; this ms reports such a strong and excellent study that it seems somewhat mealy-mouthed to say "This limitation is mitigated by.." it rather sounds like the difference between saying "I'm sorry" versus "I'm sorry you feel that way" which is no kind of apology. It could be reworded to avoid this.

Line 203: Tiny comment, but the "so" is ugly.

Line 204: make them a unique model suited to probing/exploring complex....

Appendix B

Referee: 1

Before responding point-by-point to this reviewer's comments, we wish to express our sincere gratitude for the thoughtful and constructive nature of this review. These comments have been very helpful in revising the paper, and we are grateful for the reviewer's input.

Comments to the Author(s)
I think that this manuscript has substantial potential to become a very nice contribution to Proc B. However, in my opinion, it would be much better if substantially fleshed out, possibly focusing on a subset of the major components that are currently presented. Following fleshing out, it seems to me that further review wold be necessary, as the amount of detail for a reviewer to go on is currently quite minimal in some aspects.

We thank the reviewer for his kind remarks and positive evaluation of the paper. We agreed that the paper would benefit from some restructuring and have done so in the revision. We have followed the reviewer's general advice to provide more detail regarding the central elements of the paper and have removed some analyses that potentially distracted from the main ideas.

Major

(1) This paper is so brief that it is very difficult to know what context the authors intend for each component analysis - particularly, what patterns might be expected, and which, if any, might be surprising. To a similar extent, the manuscript is quite vague on what is to be made of different results. While I don't presume to know exactly the history of the paper, I think the authors would have to forgive the reader for wondering if this is a hasty re-submission following an initial try at a journal requiring a much shorter format. While I recognise (and like) that PSRB is flexible in the format of initial submissions, I feel that, with the many related results given in this manuscript, the authors have failed to make good use of the space afforded by PSRB lay out the background and interpretation of the various components of their paper. I think that the "among-breed heritability" and the GWAS components could be more fully described, for sure, but the additional studies beyond these two most developed components have such minimal treatment that the really feel like add-ons. I suspect that a fully-developed treatment of just the heritability-like analyses plus the GWAS could be a the basis fo a nice paper.

We agree that the original structure of the paper made it difficult to fully grasp some components of the work. The original submission was formatted to stay close to the 6-page / 5100 word standard article format for PSRB. The length for the original submission was 5,180 words, estimated to be 8 printed pages including figures, using the PSRB page estimation tool. However, we have revised the MS to use a more traditional format, presenting methodological details before the results, and fleshing out several areas suggested by the reviewer, as described below. In order to make room for additional detail, we have also eliminated some aspects of the paper which were less central to the

manuscript (see below). The revised manuscript is ~6,200 words and estimated to be 9 printed pages.

(2) I suggest that the use of the word “heritability” should be avoided, and also, the relationship between what is currently referred to as “heritability” at the among-breed level to heritability within breeds should be carefully articulated. Across biology, “heritability” is used here and there to denote various aspects of the degree of genetic control of phenotype, generating lots of potential confusion. In evolutionary quantitative genetics, “heritability” refers to the proportion of phenotypic variation that is attributable to (additive) genetic effects. This proportion is what is calculated here at the among breed level, but it is the level that I think requires reconsideration. What the proportion heritability determines is the proportional response to selection (i.e., how much of the selection differential is represented by the response to selection) when selection is applied to some population to which selection might plausibly be applied. It is a property of some (essentially random mating) population at a given time, in a given environment. As such, heritability — in the context of the evolutionary QG principles that gives heritability its meaning — is the within-breed proportion of genetic variance. An analogous proportion calculated at some other level of biological organisation is fundamentally a different quantity. The calculation that has been performed is closely analogous (a proportion of variance), but I think that the biological meaning is so different that (a) a different word ought to be used, and (b) there paper needs in some way to reflect this difference.

This issue of wording and the theory underlying the notion of heritability comes back to my point (1) as well — what did the authors expect, and why, for the magnitude of the proportion of among-breed phenotypic differences that are genetically based? What did they expect for the difference between the existing within-breed heritability estimates and the new across-breed estimates of the proportion of genetic effects, and how is the difference to be interpreted? Are they to be thought of as the same thing (I think not, but to some extent the authors seem to), but the current one is somehow a better, bigger number? If they are not to be understood as representing the same thing, then what is to be understood from their comparison?

We agree that discussing heritability at different levels of biological organization can present some confusing issues. We debated what best to call the measure used in this paper and considered the appropriateness of several terms and analytical approaches. As the reviewer notes, heritability typically refers to the proportion of phenotypic variance attributable to additive genetic variation within a population. At broader levels of biological organization (e.g. between species) the most conceptually related (but certainly distinct) measure might be phylogenetic signal. However, phylogenetic signal relies on estimates of the amount shared and non-shared evolutionary history from a phylogenetic tree, as well as an associated evolutionary model regarding the accumulation of trait variance (e.g. Brownian motion). Because dog breeds were created predominantly through repeated crosses between different forms, rather than evolution along a bifurcating tree, we concluded that phylogenetic approaches, and the related terminology, were inappropriate for our dataset. In the revision we have chosen to use the term ‘among-breed heritability’ (borrowing from the reviewer’s description of this level of analysis), to

disambiguate our cross-breed measure from typical heritability estimates in a randomly interbreeding population. We thank the review for helping us to present this more clearly.

We have opted to retain the general concept of heritability (but designating it as 'among-breed' heritability) primarily because 1) The calculation of our h^2 measure is mathematically the same as it would be in a within-breed heritability study incorporating an IBS matrix to capture relatedness (e.g. Ilska et al, *Genetics*, 2017); and 2) All dog breeds are interfertile, and in a biological sense are members of a *potentially* interbreeding population, in the absence of human intervention. Indeed, new breeds of dogs are created through the crossing of existing breeds (e.g. Labradoodles, Yorkipoos), and estimation of heritability among breeds would likely inform the response to selection in these contexts (e.g., considering all dogs as a base population one could calculate a selection differential given parents of breeds X and Y, and infer a selection response using among-breed heritability estimates). The revised paper makes clear that the two uses of the heritability concept have important differences, but also share a common foundation related to the proportion of variance attributable to genetic factors in a given study population (see lines 73-91, 215-219, 325-333).

(3) Having laid out my points (1) and (2), I can try to articulate my point (3), which is really a hybrid of the first two points. If more space is taken to outline the background and some sorts of expectations, what would those expectations look like? On the main issue the magnitude of the heritability-like proportion calculated at the among breed level, are the qualitative findings expected or unexpected? I think they are exactly as we would expect. Regardless of the exact level of heritability, in the strict sense, so long as h^2 isn't trivial, then sustained diversifying selection, as has inarguably happened in dogs, will eventually generate large genetically-based differences among breeds. Taken long enough (and surely in dogs it has progressed long enough, considering the various reasons why it is obvious anyway that there are massive genetic differences, e.g., two breeds both raised in the common environment my kitchen look and in many ways act totally different from each other but perfectly typical for their breed, hybrids are often very much intermediate in many characters including temperament, etc…), then the proportion of among breed variation that is genetic will inevitably exceed the proportion of within-breed variation that is genetic. Similarly, if there are genetic differences that are picked out by the linear modelling analysis, then, power issues notwithstanding, one would hope that at least some of the underlying loci, which must necessarily exist, could be detected (and this seems to be the case). I do not necessarily think that the fact that everything might be neatly as expected is a bad thing — so much evolutionary QG gets massively and inconveniently complicated, that a neat study laying out what we should expect in a system like dogs, and showing that things do indeed fit together, could be a lovely thing.

We thank the reviewer for this thoughtful comment and agree that our results are very much in line with what one would expect given the recent history of diversifying selection in dogs. In the revision, we more clearly set up these expectations in the introduction (lines 73-98) and conclude that dogs provide a relatively 'clean' system in which to evaluate

these processes at a genetic level, with the current results being largely in line with expectations (lines 328-336).

(4) The fleshing out that I'm suggesting may not necessarily be entirely presentational. I've been trying to think a bit about what more carefully drawn links to underlying theory and expectations might look like. While the current discussion give little explicit guidance on what to think about the high degree of genetic determination of breed differences, it seems that the intention is that the reader sees this as a signature of responses to diversifying selection. However, drift alone would generate variation among breed means in proportion to the degree of genetic differentiation among breeds. This is exactly the signal that is quantified statistically in the GRM-mixed model analyses. The effects of selection are actually represented in whatever genetic differences exist, over and above those that would arise under drift. I suspect, given how recently the breeds have diverged, and despite the likely bottlenecks in the development of many breeds, that the diversification far exceeds what would happen under drift - but some kind of F_ST vs Q_ST approach, or better yet, calculations (based on admittedly a certain amount of guesswork e.g., number of generations, effective population sizes, etc.; even quite rough back-of-the-envelop work could be really interesting), would be necessary to relate the mixed model results to the degree to which the genetic basis of among-breed differences can be related to different evolutionary processes.

These are wonderful suggestions and we have now incorporated a Qst-Fst analysis in lines 160-165, 225-226, and Figure S1. This approach demonstrates that breed differences in dog behaviors are likely due to strong selection and not genetic drift due to recent demographic history of breeds. Pst, a proxy for Qst, ranges from ~0.6-0.95, while genome-wide Fst is approximately 3-fold smaller (0.255).

(5) While the methods do cover all analyses, I think that they could stand to be fleshed out quite a lot too. Many analyses are described via the statistical packages that were employed, and in some cases the names of the options used in those packages. In five or ten years, many of these packages will be obsolete and people will have forgotten (and in many cases will never have known in the first place) what these packages and options did. I think that descriptions need to be given much more of what the actual biostatistics were. For e.g., (this is not exhaustive) (i) what measure of genomic relatedness was used to make the genomic relatedness matrix (there are multiple potential options, and even if your choice was quite standard, what is standard now may not be in five years time)? (ii) What were the breeds of the four dogs in the gene expression study (and does it matter, or more importantly, as it doesn't seem to matter massively to the authors, why doesn't it?)? (iii) What is the reader supposed to understand from the one bit of detail about "raw questionnaire item scores, rather than factor scores" (line 264) - this seems to assume knowledge about the data that the reader does not have?, (iv) of the 25 individuals per breed that were subsampled for each mixed model run, were their phenotypes averaged, or used as separate measures? If the former, say so, if the latter, then there must be an additional variance component in the model (giving an equation for the mixed model would be a great way to help flesh out the description and solve this kind of ambiguity).

We thank the reviewer for pointing out methodological components that needed additional description, which we have provided in the revision. In general, we have added methodological details throughout, but provide responses to the reviewer's specific queries below:

(i) The genomic-relatedness matrix was generated as an identity-by-state matrix. Off diagonal elements in this matrix reflect the proportion of SNPs from the array that (on average) are identical between pairwise combinations of breeds. This matrix was generated in PLINK using v1.9's default options for generating an IBS matrix (plink --file datafile --cluster --matrix). From this matrix, we then generated our breed-averaged IBS matrix by calculating the mean IBS within breeds and between pairs of breeds. We now provide additional details on the relatedness matrix used, with an example, along with a model equation in the supplemental material.

(ii) Two beagles and two mixed breed dogs were used in the expression study. While gene expression likely varies between breeds, broad patterns of gene expression tend to be highly conserved (as evidenced by highly similar tissue-specific expression profiles in humans and dogs), and we have noted these details and justification in the revision.

(iii) We have added much more detail regarding the C-BARQ, its factors, and the behaviors they reflect, including a new table (Table 1). We have removed the analysis that used the individual items rather than factor scores, but now better explain the details of this instrument in the methods section (lines 124-133).

(iv) We agree that it is helpful to see a formula for the mixed model as well as some of its components (how we generated the K matrix). Our heritability analyses did incorporate individual-level data and details about these models is now shown in the supplemental material.

(6) There seems - but it certainly could be my misunderstanding - to be an unnecessary convolution in the estimation of the heritability-like parameter. I gather from 246-249 that many analyses were done, using just a few records per breed. The reason for this is not given, but from first principles of mixed modelling, this should not be necessary. A mixed model should be able to use the raw individual-level data, and partition that into a residual variance (differences among individuals, within breeds), and breed-level variance components that are and are not associated with the genetic differences among breeds; the heritability-like ratio constructed from these latter two components would be the proportion of breed differences that is due to genetic differences among breeds. This would be a lot neater, and would make full use of all the data in a single analysis. If this analysis was too slow because of the many individual records, then a mean and SE for each breed could be calculated, and these could be used as meta-analytic data in MCMCglmm (and some REML softwares could also accommodate this alternative parameterisation), wherein there could be a GRM-linked variance component, and the residual variance would represent among-breed, non-GRM related, differences among breeds.

We apologize that this was not presented clearly. For the heritability models, we used individual-level data in each model, but performed the analysis across resamples, each iteration drawing N = 25 from each breed. This was done in order to reduce computational demands and assess a distribution of results across subsamples (see lines 150-152). We appreciate the reviewer's suggestion of a meta-analytic approach with MCMCglmm, and will consider this approach in future work.

(Signed; n.b. all my reviews are signed) Michael Morrissey

Specific

- line 54: two points. I would consider dropping the "compared to humans" - at least for PRSB, I would think that the biological principles, irrespective of species, are of interest. I would delete the word "uniquely" (also l.203-204) - I don't think it is true, an whether it is true is probably not relevant.

We have implemented both suggestions.

- line 58: how a "limited genetic toolkit"? What is a "genetic toolkit"? I feel this kind of phrasing is common, but just muddies things, as there is really no specific meaning. I'd remove.

We have reworded this sentence to be more precise.

- l.82 "control for differences among breeds" this is the kind of phrasing that works best if we are trying to control for some confounding variable. Actually, the relatedness among breeds is they key source of information here, not some nuisance!

Indeed! We have removed this phrase, which belongs in the GWAS analyses, where we really are attempting to control for population structure when assessing associations between SNPs and behavior.

- l.87: in my opinion, it would be much better to give means and standard errors (or better yet, differences, and SEs of the differences) rather than test statistics and p values without any notion of biological effects, or uncertainty in those effects. This applies throughout.

We have added the mean difference and 95% CI.

- l.93: "cultivation" seems a strange word

We have changed the wording here.

- l.94/95. Yes, there is undoubtedly a correspondence between the genetic and phenotypic clustering. But there is no sense in which they recapitulate one another with the "high accuracy" that you describe. The ability to predict one from the other is very vague. Also, this might be one

of the best of many opportunities to reinforce my main point (1). The many potentially interesting facets of the work that went into figure 1C are represented by only this one statement (which I believe to be superficial and over-stated). I think there is a potentially a lot of value here that is greatly suffering from the extremely abbreviated presentation.

We agree that the correspondence is not overwhelming, and that the limited text regarding this figure is challenging for the reader. In order to flesh out other, more central aspects of the paper, we have removed this panel from Figure 1, and dropped the clustering analyses which were not critical for the 'main story'.

l.119: Bonferroni doesn't (classically) change the p value, but rather the critical p value (a.k.a. the alpha value). It seem unnecessarily confusing to state it this way, as though the p values themselves have somehow been corrected.

We agree this was potentially confusing and now simply state that we used Bonferroni correction.

l.26-128. Calculating the variance associated with top SNPs is actually a really difficult thing, because, on average, the *estimated* effect sizes of the top SNPs will surely be (on average) overestimates of the *true* effects of the top SNPs. This is the GWAS analogue of the Beavis effect in QTL analysis. I'm not sure if your calculations here try to control in some way for this effect (i.e., if it is anything more than a sum over loci of 2pq\alpha), or even what your calculations are. It is another case of much more methodological detail being necessary before a real evaluation can be made (ether by a reviewer, or more importantly, a reader).

This is an important point regarding possible inflation of effect sizes in GWAS studies. However, we expect that the variance explained by SNPs with the lowest p values will be similarly inflated in within-breed GWAS studies. Thus, while the actual magnitude of the effect may be inflated, we expect that our finding of more variance explained across than within breeds is robust to this source of error. We have clarified this important issue in the revision (lines 259-263).

l.130-134, l.147/8. The supplemental stuff referred to here seems to me to actually be the important "meat" of a major component of this manuscript. I think it could be much more developed - maybe even into a manuscript of its own! (n.b., I hope you won't think I'm advocating for minimal publishable units here — I think there really could be multiple meaty papers here!)

We agree that the enrichment analyses are a meaty part of the paper deserving of some unpacking. In our view, these analyses are inextricably linked to the GWAS, because the aim of our GWAS is less to identify SNPs with causal effects (a challenge with limited genomic coverage in array-based studies) and more to identify regions of the genome implicated in behavioral differences, and possible functional roles for the associated genes. Because we take a meta-analytic approach, aggregating effects across SNPs to identify promising genes, the enrichment analyses serve an important role in interpreting

our results at a broader scale (e.g. are the variants we find in genes that plausibly contribute to behavioral differences through expression in the brain or involvement in biological processes relevant to behavior?). However, we agree that this section needed additional fleshing out, and have removed other elements of the paper to make room for more detailed treatment of the enrichment analyses.

From l.169: way to little detail here, in my opinion.

We agree and have unpacked the sections on enrichment analyses in the methods, results and discussion.

l.194: I would not describe this as a "limitation". The information needed to study differences among breeds, which is something entirely worth studying, is well-provided by the data to hand!

Thank you, and we have revised the wording so as not to frame this as a limitation.

Reviewer 2

Two questions are posed and answered in this ms: 1) to what extent are behavioural traits heritable in dogs? 2) If heritable, what is the genetic architecture of the traits? Are they largely polygenic?

This paper is a corker. The questions are worth asking, the research design provides a clever way to probe the questions, absent individual level behavioural and genetic data on the same dogs. The interpretation of the results sticks well to the data without over-reaching. It's an ambitious piece of work that looks well executed and mostly very well written. It will make an excellent contribution to the literature.

We thank the reviewer for his or her kind words and enthusiasm about the paper.

I have some comments for the Authors to consider:

Spell out the essential design of a little more clearly. I thought that since this paper does not use the standard behavioural genetic designs to explore answers to the key question about trait heritability, it would be helpful to further elaborate the design. The design is so thoroughly well-known to the Authors; it is hard to escape the curse of knowledge. For example, readers may wrongly think that there is a breed specific reference genome for each of the 101 breeds in the study. It wouldn't take much work by the Authors to allow readers to follow a little more easily what was done.

Thank you - we have restructured our description of the design and added additional detail regarding the genotyping platform and incorporation of these data.

The Results begin with Methods, I'd consider starting with " We found that... and moving the part before it.

We agree that a format change was beneficial, and now present all methods in a dedicated methods section prior to presenting the results.

The 14 behavioural traits.
What I wanted to know for each of these was: an example showing how the trait scores are derived and an index of the reliability of each trait. The heritability of mush is less interesting than the heritability of a well-measured trait so it's important to know something more about the measures. In line 90 you write about h2 estimates - is h2 correlated with reliability? Are more reliably measured traits more heritable? Many readers will not know C-BARQ - I assume some traits are more reliably assessed than others. Could this be given in the supplementary materials so that readers do not have to first research C-BARQ?

We agree that additional details about the C-BARQ are important to include and we now present this instrument (including validity metrics) with much more detail in the methods section as well as a table describing properties of the 14 factors (Table 1).

Fig1 B and C
I printed these out and also looked at them on screen so I could increase the size of the text. I couldn't read them either way. Consider a note to say that a larger version is provided in Supplementary materials?

We have removed panel C from Figure 1, which allows panel B to be presented at a larger size. We will also be happy to provide higher-quality images as needed for publication.

Fig2 A
The colour index is a little too fine grained- one suggestion is that a Manhattan plot for selected traits is presented in main text with a note that Manhattan plots for other traits are in Supplementary. Just an idea.

Thank you for this suggestion. We have added a supplemental figure that includes the Manhattan plots separately for each trait.

The trait 'Energy'. I don't know what this means. Please can it be explained? Line 153. Assertion needs a citation (genes associated with intelligence and information processing).

As noted in response to the reviewer's comment above, we have added a more detailed description of all behavioral traits measured by the C-BARQ (Table 1).

Line 86-87: worth mentioning range-restriction - to explain the difference in estimates? I'd lose the "identified" it's redundant here.

We have dropped the 'identified' and now address the differences between within- and among-breed heritability estimates much more explicitly (see responses to Reviewer 1; and Lines 124-133, 211-215).

Line 191: BG is aset of methods, so I'm unsure what is meant by variant being "implicated in human BG" do you mean that there may be the same variants may also be associated with tameness/ aggression in humans? If so, perhaps make clearer and give a citation.

We have clarified that variants in same genes have been associated with similar phenotypes in humans, and now refer to the table which includes the associated references.

194-197 Limitations; this ms reports such a strong and excellent study that it seems somewhat mealy-mouthed to say "This limitation is mitigated by.." it rather sounds like the difference between saying "I'm sorry" versus "I'm sorry you feel that way" which is no kind of apology. It could be reworded to avoid this.

Thank you. Reviewer 1 had a similar concern and we have revised the wording in this section accordingly.

Line 203: Tiny comment, but the "so" is ugly.

We have dropped the ‘so’.

Line 204: make them a unique model suited to probing/exploring complex....

We have changed the wording in this sentence as noted in our replies to Reviewer 1.

Appendix C

Editor comments:

there were concerns shared in the confidential comments that your use of the term "heritability" is somewhat unconventional and so it would be helpful if you justified your terminology in the introduction somewhere

We have included text in the introduction to justify and clarify our use of this term, as follows:

When considering variance across breeds, we can make use of the heritability concept while acknowledging some potentially important differences from studies of heritability in randomly interbreeding populations….To disambiguate within- from across-breed measures of heritability (acknowledging fundamental differences in the reference populations for these studies), we use the term 'among-breed heritability'

Referee 2:

"For example, although breed differences in behavior are well documented (reviewed in 3), it remains unknown to what extent these differences are attributable to genetic relatedness among breeds. "

"First, we hypothesized that if diversifying selection in dogs has led to genetically-based breed differences in behavior, then much of the behavioral diversity among breeds should be attributable to genetic similarity among breeds. "

In both the above sentences, the reader has to work hard. In standard QG reports, the focus is on finding the extent to which phenotypic similarity is predicted by genetic similarity. Seeking diversity to explain similarity is hard work to untangle - could this be worded more intelligibly?

Thank you for helping us word this more clearly. We have revised these sentences as follows:

"Despite rapid progress in these areas, we still know little about the biological bases of breed differences in behavior (reviewed in 3). For instance, the extent to which phenotypic similarities are predicted by genetic relatedness among breeds remains unknown".

"First, we hypothesized that if diversifying selection in dogs has led to genetically-based breed differences in behavior, then phenotypic similarity among breeds should be attributable to genetic similarity among breeds"

Line 92
Later you say "Second, if breed differences in behavior are largely attributable to genetic differences among breeds, ..." this has the virtue of both nouns (breed differences, genetic differences) pointing in the same direction.

We have rephrased this sentence as follows:

"Second, if breed differences in behavior are highly heritable, we expected that models predicting breed-average behavioral scores as a function of breed-average allele frequency would identify loci contributing to phenotypic variance."

Lind 78-80
When considering variance across breeds, we can make use heritability while acknowledging some potentially important differences from studies of heritability in randomly interbreeding populations

Typo 'of' missing in action?

Thank you – this typo has been corrected.

Referee 3:

Comments to the Author(s)
I think this is a very good paper -- but the Abstract somewhat undersells the findings. I would find stronger terms than "provide insight" and "suggest that dogs provide a powerful model." This makes it sound like the papers main findings are methodological, whereas there are real findings here that people care about: Namely that we now have estimates of among-breed heritabilities of certain personality types; they are quite high. Further that the specific genes responsible for typical behavioral differences between breeds of dogs have been identified.

We thank the reviewer for his or her positive appraisal of our work. As suggested, we have changed the final sentence in the abstract to emphasize the importance of our findings in dogs, rather than the methodological virtues of working with this system. The revised abstract is below:

Variation across dog breeds presents a unique opportunity to investigate the evolution and biological basis of complex behavioral traits. We integrated behavioral data from more than 14,000 dogs from 101 breeds with breed-averaged genotypic data (N = 5,697 dogs) from over 100,000 loci in the dog genome. We found high levels of among-breed heritability for 14 behavioral traits (the proportion of trait variance attributable to genetic similarity among breeds). We next identified 131 single nucleotide polymorphisms associated with breed differences in behavior, which were found in genes that are highly expressed in the brain and enriched for neurobiological functions and developmental processes, suggesting that they may be functionally associated with behavioral differences. Our results shed light on the heritability and genetic architecture of complex behavioral traits and identify dogs as a powerful model in which to address these questions.