Dear Professor Craig E. Franklin and Editorial Team,

We would like to thank you and the reviewers for providing extremely constructive feedback on our manuscript entitled: “**Metabolic consequences of sex-reversal in two lizard species: a test of the like genotype and like phenotype hypotheses”.** We are glad the Editor and Reviewers found the paper of general interest for the readers of the *Journal of Experimental Biology*.

We have now carefully considered all the comments and revised our main manuscript and supplementary materials to deal with the comments. Below we provide a line-by-line response (in ‘blue’) to each of the comments raised by the Editor and two reviewers (in ‘black’). Where relevant, we have pasted the section of our manuscript we have edited to provide clarity to what we have done to address comments.

We believe that our revised manuscript is significantly improved. We hope that you now find it suitable for publication in the *Journal of Experimental Biology*.

Sincerely,

Kristoffer Wild (on behalf of all authors)

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Editor -**

As you will see, the reviewers found merit in your study but have raised substantial criticisms that prevent me from accepting the manuscript at this time. Both reviewers felt that there was too much emphasis on fitness, which detracted from the hypothesis you tested. They recommend removing the fitness discussion and rewriting the Discussion.

The reviewers are willing to see a revised manuscript. If you think that you can deal satisfactorily with the criticisms on revision, I would be pleased to see a revised manuscript. We would then return it to the reviewers.

In revising your manuscript, please take into account the formatting instructions detailed below (we accept over 95% of revision submissions and therefore hope you won’t mind any extra work involved in reformatting your manuscript at this point).

Please ensure that you clearly highlight all changes made in the revised manuscript. Please avoid using 'Track changes' in Word files as these are lost in PDF conversion.

I would be grateful if you would also list how you have dealt with the points raised by the reviewers in the 'Response to Reviewers' box. Please attend to all of the reviewers’ comments. If you do not agree with any of their criticisms or suggestions please explain clearly why this is so.

I look forward to receiving your revised manuscript.  
**Response**: We thank the editor for their time and management on our manuscript.

Your insightful comments and constructive feedback have been immensely helpful in improving the quality of our paper. We have carefully considered your and the reviewers’ suggestions, and overall, we do agree with the criticism of limiting the manuscript’s focus on fitness. Therefore, we have made extensive revisions to remove our focus on fitness within the manuscript as a whole. Specifically, you will see we made substantial edits within the discussion to focus more on the explanation of sex differences in relation to the Like Phenotype/Like Genotype framework rather than fitness differences.

You will find that we have also formatted the manuscript to fit journal guidelines, and we hope it is considered for publication. Once again, we appreciate your time and effort in reviewing our paper. Your comments have been invaluable in refining our work. We hope that you will find our revised manuscript satisfactory.

**Reviewers' Comments to Author**:

**Referee: 1**

Comments to the Author(s)

This MS diagnoses differential phenotype investments (metabolic investments, growth and survival) among genotypic males and females and sex reversed individuals in two species of lizard subject to population-specific environmental sex reversal in Australia and that differ in genotypic heterogamety. The authors test the like-phenotype like-genotype hypothesis experimentally using appropriate experimental design and response variable quantification in a complex and interesting system. Alternative hypotheses are structured and well explained allowing conclusions to easily be reached. In my opinion the actual hypothesis   
test (like P like G) in this system is more than enough to warrant publication and is novel and significant:

1)The tests of metabolic profile, survival and growth do not remotely serve as a fitness proxy. This is the same problem that Warner and Shine (2008) [responded to by Pezaro et al. 2017] ran into. Survival and growth, nor metabolic rate, can explain selection-based activity on sex-reversal mechanisms within or among populations, NOR is this discussion relevant to the like P like G hypothesis test!

If individuals were raised to F2 and fitness could be quantified, then discussion of fitness would be warranted. Further, and in contradiction to myself, results show that sex reversed individuals had lower survivorship -- an observation that may be meaningful if we incorrectly assume that all surviving individuals had surviving F2 hatchlings, and thus higher fitness. The phenotypic profile of these sex-reversed individuals between genotype and sex phenotype -- the interplay between molecular code and specific endocrine organization -- is the interesting test, and fitness discussion at all only burdens this manuscript and distracts the reader from the interesting hypothesis test. If the environment produces sex reversed individuals in these populations and those individuals breed, then why does an adaptive value of a sex reversed phenotype become the immediate thirst to be quenched? I suggest removing any discussion of fitness from the MS pertaining to your test and results.

**Response**: We appreciate that the reviewer values the design of our “Like Phenotype/Like Genotype” system that is presented in the manuscript. We do agree with the reviewer’s main criticism of reframing from the fitness focus and rather concentrating our results within the framework of the hypothesis (like P/Like G) being tested. You will see that we have reworded portions of our abstract and introduction to reflect this criticism. We have also rewritten portions of our discussion to reframe the previous focus on fitness differences. We have introduced how phenotypic sex differences could be driven by other mechanisms such as molecular code or hormonal factors and highlighted areas of lack of understanding of how sex differences in life history traits may be connected to the frequency of sex-reversal in the wild.

2) Can the authors explain why bayesian statistics were implemented? As it seems, no prior was implemented and the data were collected and in hand (i.e. metabolic, rate, growth, survival), so why is posterior probability required? I agree a mixed effects model is used correctly.

**Response**: We used the default prior settings for mixed-effects models within brms which uses a weakly informative prior distribution for the intercept and slope parameters. For all population-level (i.e. Fixed effects), the default prior for the intercept is a normal distribution with a mean 0 and standard deviation 10. The default prior for the shape parameter of the intercept was a Student-t distribution with mean 0, scale 2.5, and 3 degrees of freedom. The default prior for residuals (sigma) was a Student-t distribution with mean 0, scale 2.5, and 3 degrees of freedom. The Cholesky factor was used as the default prior for correlations between random effects. Together this provides a weakly informative prior with a wide range of possible values (Bürkner, 2017). The prior distributions within brms are designed to be vague enough to allow the data to drive the inference, but informative enough to help scale the estimation of the parameters. Given these parameters, we feel this approach is conservative in its estimations in comparison to a frequentist approach (see Dennis 1996; Dennis 2004; Ellison 2004 – for good overviews of the arguments of Bayesian and frequentist inference). Additionally, the ultimate strength of using a Bayes approach for our experiment is that we were easily able to extract posterior distributions to test specific hypotheses within the “Like-genotype/Like-phenotype” framework. We agree with the reviewer’s point to address the reason and clarity of this method. So we have provided a short explanation of why the Bayesian approach was used in the methods section:

*“Using Bayes’ theorem allowed a more intuitive and flexible framework for interpreting the results of hypothesis testing in comparison to a frequentist approach. For example, each Bayes model allowed for a direct measure of the evidence for or against each hypothesis within the Like Genotype (genotype - sex-reversed) or Like Phenotype (phenotype - sex-reversed) framework, which was expressed as a probability or posterior distribution. In contrast, frequentist hypothesis testing would have only focused on the probability of the data given the hypothesis being tested, which can be more difficult to interpret and can lead to issues with p-value interpretation and multiple testing (Ellison 2004; Bürkner, 2017).”*

We have made it more clear in the methods section that default priors were used for all Bayesian models. Specifically, we have provided a citation of the prior default setting for brms models and added within the supplementary material the prior response to the reviewer above.

***In manuscript:***

*“For all Bayesian models, the default priors setting was used, and a description of the default prior settings for mixed-effects models within brms can be found in Bürkner (2017)*

***In supplementary materials:***

*“Default priors for all Bayesian models were used. For all population-level (i.e. Fixed effects), the default prior for the intercept is a normal distribution with a mean 0 and standard deviation 10. The default prior for the shape parameter of the intercept was a Student-t distribution with mean 0, scale 2.5, and 3 degrees of freedom. The default prior for residuals (sigma) was a Student-t distribution with mean 0, scale 2.5, and 3 degrees of freedom. The Cholesky factor was used as the default prior for correlations between random effects.”*

**Referee: 2**

Comments to the Author(s)

This manuscript seeks to determine whether there exists physiological variation between matched and mismatched genetic and phenotypic sexes in two species of lizards exhibiting temperature dependent sex reversal. The authors obtained hatchlings for two species, maintained them in the laboratory and measured metabolic rate (SMR), growth rate, and in-lab survival over several weeks.

Overall the main findings were that the body mass scaling of MR differed in sex reversed individuals relative to the matched phenotypic sex, but that the patterns differed between the two species. Growth and survival did not differ among groups.

This was an interesting study and the topic of sex reversal in response to environmental cues is fascinating. My over-arching criticism/concern with this paper is the over emphasis on the fitness relevance of SMR. Energy turnover and allocation is very likely of immense importance and has fitness consequences, but this does not necessarily mean that, on an individual level, variation in SMR has any correlation with fitness. SMR is but one piece of the energetic pie (but admittedly an important one). But, the relatively few studies that have sought to identify a clear link at the population level between SMR and fitness have not clearly demonstrated a link. Therefore, the premise that this paper is based on is not terribly well supported. In this study, was SMR correlated with individual variation in growth or survival? This information would be helpful, even though all animals were raised during the experiment indoors under lab conditions.

**Response**: This is an excellent point. Thanks for raising it. We have gone through the entirety of the manuscript to revise the overemphasis on the fitness relevance of SMR. Also see our response to reviewer 1. Overall we have focused our efforts on discussing sex-differences within the framework of the like-genotype/like-phenotype hypothesis and have provided alternative explanations within the discussion section.

As for point 2, we do feel that testing if SMR had any correlation with growth or survival is an interesting point, but we feel it is now out of the scope of like-genotype/like-phenotype framework.

Below I make some additional comments that hopefully help to improve the paper.

Minor comments

1. L77-79: Please cite some empirical studies clearly demonstrating a correlation between metabolic rate and fitness. This is a pretty loose and likely very indirect correlation

**Response**: We agree with this point and have removed the statement linking energy expenditure to fitness. We have also provided an additional citation of a review that links RMR to growth and survival (Burton et al., 2011 see Table 1 for additional examples of empirical studies showing relationships between RMR and fitness-related traits). The re-written sentence follows:

*“In both empirical and theoretical studies, estimates for metabolism have shown to be inextricably linked to individual patterns of growth, reproduction and survival (Peterson et al., 1999; Burton et al. 2011; White et al., 2022).”*

2. L118-119: This statement about metabolism is extremely general. In your study you measured SMR..."metabolism" is too general here. SMR is only one component of energetics.

**Response**: This is a great point. We have rewritten this sentence to fit the reviewer's point and accurately address where there is a hole in the literature.

*“To date, no studies have explored any components of energetics (i.e. metabolism, growth, maintenance) and if consequences associated with sex-reversal in any other vertebrate, even though sex-specific strategies of energy allocation have been documented between males and females (Geffroy, 2022; Somjee et al., 2022).”*

3. L186: Did you use tissue samples for DNA extraction or blood samples? Both are mentioned here.

**Response**: We used tissue samples. Blood has been removed, thank you for bringing this to our attention.

4. L286-288: For growth rate analyses, what was the random effect used in the mixed models? My understanding was that each lizard was represented by a single measurement. So, was clutch of origin the random effect? Please clarify.

**Response**: Great catch. We have clarified that growth rate models were compared across sex class using Bayesian linear model.

*“Differences in growth rates were compared across sex class using Bayesian linear models. Growth rate of SVL and mass were analysed as a function of initial size (or mass) measurements, sex class and their interaction.”*

5. L320-322: Maybe I'm misunderstanding this statement, but do you mean to say that the mass scaling of MR changes with size? Seems odd...

**Response**: Yes, we misspoke here. The sentence has now been corrected to accurately reflect that the mass-specific metabolic rate varied across sex. We then used the follow up sentence to describe the scaling pattern:

*“Sex-reversed female P. vitticeps (femaleSR ZZ) had a mass-specific metabolic rate that was overall higher than their genotypic counterparts (male ZZ - femaleSR ZZ; pMCMC < 0.01), but lower than their phenotypic counterparts (female ZW - femaleSR ZZ; pMCMC = 0.04; Table 3). As female ZZ got larger, the mass scaling relationship of metabolism was more like ZZmales than ZW females (Fig. 2D)”*

6. L361-363: This seems a bit of a stretch. Do you have evidence that animals with an "energy surplus" are more aggressive or active?

**Response**: Agreed. We have reworded this paragraph to focus on the argument of how resources could explain body size differences and fecundity differences between phenotypic sex observed in laboratory settings.

*“ If all other aspects of the energy budget are the same,* *a similar sized femaleSR ZZ would have more residual energy than female ZW to allocate towards maintenance and growth after resting metabolic costs have been paid. This surplus in energy reserves for femaleSR ZZ may explain why adult femaleSR ZZ P. vitticeps are larger like males and more fecund than female ZW in laboratory conditions (Holleley et al., 2015). However, in nature, these phenotypic differences may also be a disadvantage for femaleSR ZZ in the wild because femaleSR ZZ have lower fecundity rates and there are no clear differences in morphology to female ZW (Wild et al., 2022). Different strategies of energy allocation in relation to resource conditions, i.e. ad libitum food in the laboratory vs stochastic conditions in the field, may explain previously observed differences in morphology and fecundity differences as adults.”*

7. L372-374: Is it not also possible that the lack of differences in SMR might  indicate that this "trait" is not at all under selection?

**Response**: We have added that the lack of differences in metabolic rate may be one alternative that this trait is not under selection and then discussed an alternative hypothesis:

*“One simple explanation for the lack of differences observed in metabolic rates and growth between male XY and maleSR XX B. duperreyi is there is little or no selection for sex-reversal during early development for this species.*

L410-411: But you only detected a difference in scaling, not in mean SMR. So it   
seems inappropriate to state that one group had lower energy requirements.

**Response**: This reviewer is correct. We have reworded parts of this paragraph to clarify our argument, which expands on how environments with low resources may provide an explanation of how sex-reversal distributed across their range:

*“ If lower mass-specific metabolic rate persists in femaleSR ZZ as they grow, locations that experience stochastic fluctuations in resource availability may allow these individuals to persist in low but subtle frequencies (Burton et al., 2011; Ricklefs & Wikelski, 2002). Further attention is needed to investigate how the availability of resources influences the development of life history characteristics in different sexes and if these responses are connected to the occurrence of sex reversal in natural environments.”*