Dear Professors Craig E. Franklin and Patricia Wright,

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 that you 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. Finally, we have provided the track-changed version of the manuscript and supplementary information so that you can see where changes have been made.

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 yours and the reviewers’ suggestions, and overall, we do agree with the criticism of limiting our discussion on any fitness consequences of differences in metabolism. Overall we agree with your suggestion of a rewrite of the discussion, and we have provided a pdf of track-changed version of both the main manuscript and supplementary information so that you can see where changes have been made. Specifically, we made extensive revisions to take the emphasis off fitness within the manuscript. We have 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.

As requested, we have formatted the manuscript to fit journal guidelines. 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 toning down our focus on fitness and rather concentrating our results within the framework of the hypothesis (like P/Like G) being tested.

While we do measure survival, which was why we discussed fitness, we agree that the survival data is based on a short period and small sample sizes. As such, we have reworded our abstract and introduction to remove the focus on fitness. We have also rewritten a great deal 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 genes or hormonal factors and highlighted areas we still lack an 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 preferred Bayesian modelling approaches because of their flexibility with respect to parameter estimation. It is also easier to interpret parameters (or derivations of parameters – e.g., marginalised means) probabilistically because such approaches compute posterior probabilities for each parameter in the model and the posterior distribution is easily manipulated. However, our models could be run within a frequentist framework without issues, and our results remain the same. We have justified this better in section 2.5 as follows:

“*We used Bayesian modelling approaches because of their flexibility with respect to parameter estimation. It is also easier to interpret and manipulate posterior probabilities for each parameter in the model*.”

Contrary to what Reviewer 1 says, we did in fact use priors on all our parameters as this is required, but we agree we should have been clearer that these were default priors. We also should have provided details on what exactly these were. We have now added this information to the supplemental material as below:

*“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.”*

We have also 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 settings for brms models and added a note to check the supplementary material for details.

*“Default priors (See Supplementary Material for Details) were used and 4 MCMC chains of 5000 were run with a burn in of 1000 and a thinning interval of 5 for the “brms” models*”

It is noteworthy that Bayesian approaches tend to provide conservative approximations of parameters (i.e., greater uncertainty) compared to a frequentist approaches when one has small sample sizes (see Dennis 1996; Dennis 2004; Ellison 2004 – for good overviews of the arguments of Bayesian and frequentist inference).”

**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 (see our response to reviewer 1).

We have also taken on board Reviewer 2’s excellent suggestion at looking into the relationship between metabolic rate and growth and survival with the caveat that such estimates are 1) based on small sample sizes and 2) conducted in the lab over a short period of life.

Adding metabolic rate to growth models did not suggest individual metabolism was related to growth in either species. While it could be the case that this relationship varied by ‘sex’ type we did not have a large enough sample size to explicitly test this hypothesis (see Table 1 below). We have now added these additional analyses in the main manuscript and modified the results as follows:

*“Differences in growth rates were compared across sex class using Bayesian linear models while accounting for individual mean metabolism. This allowed us to test if there was a relationship between metabolism and growth rate (mass or svl) across sex class.”*

*Table 1: BRMS Model coefficients for SVL and mass growth rate estimates across sex class and metabolism for Bassiana duperreyi. Growth rate was calculated by dividing the change in growth (SVL or mass) between the initial measurement and subsequent remeasurement by the total number of days elapsed. Due to the small size and rate of change in grams, mass was converted to centigrams (cg). Animals were remeasured between 3 and 6 months post hatch. Metabolism was estimated by the mean log O2 measurement for each individual from metabolism experiment.*

| *Growth rate* | *Covariate* | *Estimate* | *l-95% CI* | *u-95% CI* |
| --- | --- | --- | --- | --- |
| *SVL(mm/d)* | *Intercept (O2\_SexFemaleXX)* | *-4.62* | *-5.02* | *-4.21* |
| *Growth Rate (SVLmm)* | *6.20* | *-2.26* | *14.35* |
| *O2\_SexMaleSRXX* | *-0.14* | *-0.73* | *0.42* |
| *O2\_SexMaleXY* | *0.10* | *-0.44* | *0.64* |
| *Growth Rate (SVLmm):O2\_SexMaleSRXX* | *-4.07* | *-16.08* | *7.37* |
| *Growth Rate (SVLmm):O2\_SexMaleXY* | *-7.84* | *-19.83* | *4.18* |
| *mass(cg/d)* | *Intercept (O2\_SexFemaleXX)* | *-4.61* | *-4.97* | *-4.24* |
| *Growth Rate (mass cg/d)* | *1.06* | *-0.98* | *3.04* |
| *O2\_SexMaleSRXX* | *0.20* | *-0.47* | *0.87* |
| *O2\_SexMaleXY* | *0.18* | *-0.45* | *0.82* |
| *Growth Rate (mass cg/d):O2\_SexMaleSRXX* | *-2.59* | *-6.49* | *1.25* |
| *Growth Rate (mass cg/d):O2\_SexMaleXY* | *-2.13* | *-5.89* | *1.73* |

*Table 2: BRMS Model coefficients for SVL and mass growth rate estimates across sex class and metabolism Pogona vitticeps. Growth rate was calculated by dividing the change in growth (SVL or mass) between the initial measurement and subsequent remeasurement by the total number of days elapsed. Animals were remeasured between 3 and 6 months post hatch. Metabolism was estimated by the mean log O2 measurement for each individual from metabolism experiment.*

| *Growth rate* | *Covariate* | *Estimate* | *l-95% CI* | *u-95% CI* |
| --- | --- | --- | --- | --- |
| *SVL(mm/d)* | *Intercept (O2\_SexFemaleZW)* | *-2.17* | *-2.77* | *-1.57* |
| *Growth Rate (SVLmm)* | *1.47* | *-0.84* | *3.79* |
| *O2\_SexFemaleSRZZ* | *0.17* | *-0.63* | *0.95* |
| *O2\_SexMaleZZ* | *0.01* | *-0.73* | *0.74* |
| *Growth Rate (SVLmm):O2\_SexFemaleSRZZ* | *-1.51* | *-4.74* | *1.66* |
| *Growth Rate (SVLmm):O2\_SexMaleZZ* | *-0.35* | *-3.11* | *2.45* |
| *mass(g/d)* | *Intercept (O2\_SexFemaleZW)* | *-1.79* | *-2.30* | *-1.27* |
| *Growth Rate (mass g/d)* | *-0.07* | *-2.55* | *2.41* |
| *O2\_SexFemaleSRZZ* | *0.00* | *-0.77* | *0.75* |
| *O2\_SexMaleZZ* | *-0.27* | *-0.91* | *0.37* |
| *Growth Rate (mass g/d):O2\_SexFemaleSRZZ* | *-0.90* | *-4.48* | *2.78* |
| *Growth Rate (mass g/d):O2\_SexMaleZZ* | *0.92* | *-2.08* | *3.83* |

We also modelled survival (assuming a Bernoulli error distribution) as a function of metabolism, lizard mass and each of the sex categories. These models are very data hungry but supported our original survival analysis (Table 2 below). However, relying on logistic regression meant that we were also able to account for mass and metabolic rate. In each case, there were no significant effects of mass or metabolic rate on early survival, but interestingly, for *Pogona*, metabolic rate did show a positive effect on survival (albeit it was not significant – p = 0.12). *Bassiana* models performed less well because of complete separation issues – no mortality was observed in the XYm sex category (table 3 ***below*;** see Table 5 ***in manuscript***). As such, parameter estimates are large and estimated with high uncertainty. Given the small sample sizes and the focus on lab-based survival data over a short period of time, we’ve decided to stick with our original analysis in the main manuscript. If the reviewers and editor think these additional analyses are useful we can also add them into the supplement. We have provided the R code for this analysis in our git hub under “Reviewer\_analysis.R” if the editor or reviewers were interested.

*Table 3.* *BRMS Model coefficients testing for the relationship between 6 month survival and mean metabolic rate across sex class for Bassiana duperreyi and Pogona vitticeps. To fit normality mean metabolic rate was log-transformed. The mass of individuals during metabolism testing was included in the model due to the detected effect of mass on metabolic rate.*

| *Species* | *Covariate* | *Estimate* | *l-95% CI* | *u-95% CI* |
| --- | --- | --- | --- | --- |
| *B. duperreyi* | *Intercept (Sex XXf)* | *-1.21* | *-23.90* | *21.58* |
| *log(mean\_O2)* | *-0.21* | *-4.07* | *3.43* |
| *Mass* | *5.37* | *-16.25* | *27.19* |
| *sexXXm* | *-0.73* | *-3.22* | *1.54* |
| *sexXYm* | *10.88* | *-0.12* | *45.22* |
| *P. vitticeps* | *Intercept (SexZWf)* | *4.72* | *-1.22* | *10.79* |
| *log(mean\_O2)* | *1.27* | *-0.43* | *3.05* |
| *mass* | *-0.17* | *-1.15* | *0.96* |
| *sexZZf* | *-0.48* | *-1.86* | *0.87* |
| *sexZZm* | *1.46* | *-0.28* | *3.41* |

Overall, these additional analyses largely support our original results, and justify toning down our focus on fitness. As such, 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.

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 been shown to be 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 how energetic components (i.e. metabolism, growth, maintenance) are affected by sex-reversal, even though sex-specific strategies of energy allocation have been documented between phenotypic males and phenotypic 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. There were no random effects because, as correctly pointed out by Reviewer 2, we only had a single growth measurement for each lizard. Also, these were eggs from different clutches. 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). The mass scaling relationship of metabolism for femaleSR ZZ was more like ZZmales than ZW females (Fig. 2D; Table 4).”*

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 restructured the discussion to fit reviewer 1 and your point on refocusing towards the like genotype hypothesis. On this point, we have clarified our argument, which points out that under most areas of the body size distribution curve (Fig 2D) there really is no differences between SR and concordant animals, but if selection occurs on larger individuals (“bigger is better hypothesis” - Ferguson & Fox 1984; Sinervo et al.,1992; Warner & Andrews, 2002) then there would be larger FemaleSR ZZ animals. We have then provided the differences in morphology, and behaviour and discussed how different stratiges of energy allocation may effect these processes:

*Given that selection for larger hatchling lizards in the wild is common in lizards (i.e. ‘bigger is better’ hypothesis; Ferguson and Fox, 1984; Sinervo et al., 1992; Warner and Andrews, 2002), this would imply energetic differences between adult sex-reversed and concordant female P. vitticeps. As such, we predict that adult femaleSR ZZ may have more residual energy than female ZW to allocate towards storage, production, or activity after resting metabolic costs have been paid. Such surplus in energy reserves for femaleSR ZZ may explain why sub-adult (<1year) and adult femaleSR ZZ P. vitticeps are more similar to male ZZ in behaviour and morphology, including higher activity, levels of aggression, and larger body size in captivity (Holleley et al., 2015; Li et al., 2016). However, further work is needed to investigate if these different strategies of energy allocation exist and how they translate to the observed differences between phenotypic females in body mass, body size, and fecundity in wild populations of P. vitticeps (Wild et al., 2022).****”***

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 agree. We have now added that the lack of differences in metabolic rate may be also explained by this trait may have no consequence to males and provided a new alternative hypothesis that links back to our Like Genotype/Like Phenotype framework.

*“In contrast to Pogona vitticeps, B. duperreyi showed strong support for the like-phenotype hypothesis. One simple explanation for this finding is that traits linked to metabolism are of little or no consequence for males. Alternatively, traits linked to metabolism for sex-reversed males (maleSR XX) in this species may not be associated with sex chromosomes and are linked to hormonal levels relevant to the phenotypic sex”*

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 that we detected differences in scaling and no differences in energy expenditure, but this is only the case when log mass is low. The fact that we detected small-scaling differences indicates that the mean magnitude of energy difference between the sexes will depend on the size of the animals being compared, which is why we present the predicted means at different SD’s in Figure 2. To be more clear, in Figure 2D we see no differences for small and average animals, but when animals are large we see that female ZZ and female ZW are in fact significantly different. We have added the contrast comparisons to figure 2B/D and added the contrast table with estimates in the supplementary information (Table S3).

Such an effect could be potentially relevant because we expect strong mortality for small sized lizards (see ‘bigger is better hypothesis’ - Ferguson & Fox 1984; Sinervo et al.,1992; Warner & Andrews, 2002). The lack of differences between sex reversed and concordant animals means that selection will be inconsequential for sex-reversal, but, if only large animals survive and there are clear energetic differences this could be an opportunity for selection to act on sex-reversal.

We have reworded parts of the discussion to clarify our argument, which expands on how environments with low resources may provide an explanation of how sex-reversal is distributed across their range: