We thank the two anonymous reviewers for their useful comments, and you can find our answers point by point below, in red and bold writing.

Revision round #1

Decision for round #1 : ***Revision needed***

Revision needed

The manuscript presents an innovative study on climate factors and reproductive seasonality in Papionins, with well-structured content and clear objectives. Reviewer 1 finds the methods section needing improvement, particularly the statistical analysis justification, and suggests minor editorial changes. Reviewer 2 notes the title's misleading implication of causality and critiques the definition of seasonality and data heterogeneity handling. Both reviewers appreciate the study's potential but emphasize the need for clearer methodological explanations and critical discussion of data quality and terminological definitions. Addressing these issues could enhance the study's contribution to understanding primate reproductive seasonality.

*by **Cédric Sueur**, 20 Jun 2024 13:46*
Manuscript: **https://doi.org/10.1101/2024.05.01.591991**
version: 1

Review by anonymous reviewer 1, 14 Jun 2024 08:36

Title and abstract
Does the title clearly reflect the content of the article? yes
Does the abstract present the main findings of the study? Yes
Introduction
Are the research questions/hypotheses/predictions clearly presented? Yes
Does the introduction build on relevant research in the field? Yes
Materials and methods
Are the methods and analyses sufficiently detailed to allow replication by other researchers? Yes
Are the methods and statistical analyses appropriate and well described? No (see below)
Results
In the case of negative results, is there a statistical power analysis (or an adequate Bayesian analysis or equivalence testing)? I don’t know
Are the results described and interpreted correctly? Yes
Discussion
Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument? Yes
Are the conclusions adequately supported by the results (without overstating the implications of the findings)? Yes

The article presents a study analyzing the effect of climate factors seasonality on the reproductive seasonality of Papionins in Africa. The objective is to test the hypothesis that unpredictability of climate seasonality decreases reproduction seasonality which allows greater flexibility to match vegetation productivity peak with conception, lactation or weaning. The work is based on data extracted from literature and from databases (remote sensing, climate). The text is very well structured and referenced. The results are clearly exposed and in-depth discussed. I am not a specialist of this subject and not able to fully judge relevance of the discussion, but I feel the work innovative and find the approach very solid. The discussion also comprises a part making a parallel between Papionins and early humans, both having left the forests for the savannahs and sharing flexible or low reproductive seasonality. For me, only the Methods section needs improvements:

L187 to 193 The explanations must be improved. I understand the principle that the monthly vectors are summed to compute the mean rbirth and that it diminishes when the births occur all along the year. But I would think that monthly lengths would be computed using something like the monthly proportion of annual births.

The circular statistic framework is based on the count of events. Each birth event has the same weight. Indeed, each birth event is transformed in a vector, characterised by a length of 1, and by an angle (in radius) corresponding to the julian date. Given the heterogeneity of the data (we could only obtain a count of births per month in most datasets, i.e. no access to the precise birth date), we considered the birth to occur in the middle of the month. The computed r_{birth} and u_{birth} take into account N (the number of events, i.e. here births) by summing up all birth events (i.e., vectors of length 1), so that the circular metrics work as if it was based on a proportion of births.

We made some modifications in this paragraph (Lines 202-226), and hope this section is now clearer.

L313 The choice of a Poisson distribution must be justified. Otherwise, I do not understand why a Poisson distribution is used since rbirth is continuous and the Poisson distribution is discrete. I suggest the use of the beta distribution with logit link which is flexible and appropriate for random variables with continuous bounded distributions. It is not necessary to transform variables with values comprised between 0 and 1 (if there are 0, Smithson & Verkuilen 2006, Psychological Methods11, 54-71 suggest y'=(y*(N-1)+0.5)/N with N the sample size). Perhaps the use of the beta regression would improve the results

Thank you for this comment, which was useful. Beta regression indeed seems like the most appropriate models given our dataset. However, the phylogenetic package we used, MCMCglmm, did not offer the beta regression function family. In order to integrate this suggestion, we switched to another Bayesian package, brms, to run a phylogenetic beta regression. The new methods and results are shown in the main text. Our main results remain qualitatively similar.

See changes Lines 337-340, 350-362, and new Table 2 shown below:

Predictor		*posterior mean*	*estimate error*	*CI*	
				lower	*upper*
Latitude		*0.11*	*0.34*	*-0.56*	*0.77*
Habitat productivity		***0.38***	***0.18***	***0.01***	***0.73***
Magnitude of environmental seasonality		*0.03*	*0.13*	*-0.23*	*0.28*
Number or rainy seasons (2)†		*-0.22*	*0.40*	*-0.98*	*0.58*
Rainfall peak breadth		*-0.06*	*0.17*	*-0.40*	*0.27*
Magnitude of environmental unpredictability		***-0.44***	***0.13***	***-0.70***	***-0.20***
Timing of environmental unpredictability		*-0.21*	*0.14*	*-0.52*	*0.05*
Habitat type	*(Mosaic forest-savannah)*$	*-0.21*	*0.41*	*-1.05*	*0.59*
	(Tropical forest) $	*0.56*	*1.45*	*-2.3*	*3.45*

† *The reference category is 1 rainy season*

$ *The reference category is open savannah habitat*

Minor remarks

L35 change ‘the environmental measures’ to ‘environmental variations’

Done

L140 suppress the new paragraph

Done

L212 change ‘In addition’ to ‘Yet’

Done

L349 All along the Results and the Discussion sections, I would prefer the use of the scientific names rather than the common names since those ones are not given in the Tables and in the Figures

Done

Table 1: give the signification of Krain

Done, written at the footnote of Table 1

Figure 4 Perhaps the use of abbreviations for the sites would improve clarity

We agree that Fig. 4 is 'heavy' to digest, however we do not feel that abbreviations of population names would make the figure or the text easier to understand. In fact, we instead think that abbreviations often make articles harder to follow and we have therefore taken the freedom of keeping the figure as it was. We are hoping that our decision is understandable, but remain open to other suggestions.

Review by anonymous reviewer 2, 19 Jun 2024 13:13

Title and abstract
Does the title clearly reflect the content of the article? [ ] Yes, [**X**] No (**The title suggests an analysis of causal relationships, but it is more of a correlative analysis**), [ ] I don't know
Does the abstract present the main findings of the study? [**X**] Yes, [ ] No (please explain), [ ] I don't know

Introduction
Are the research questions/hypotheses/predictions clearly presented? [**X**] Yes, [ ] No (please explain), [ ] I don't know
Does the introduction build on relevant research in the field? [**X**] Yes, [ ] No (please explain), [ ] I don't know

Materials and methods
Are the methods and analyses sufficiently detailed to allow replication by other researchers? [**X**] Yes, [ ] No (please explain), [ ] I don't know
Are the methods and statistical analyses appropriate and well described? [ ] Yes, [**X**] No (**see review as text**), [ ] I don't know

Results
In the case of negative results, is there a statistical power analysis (or an adequate Bayesian analysis or equivalence testing)? [ ] Yes, [ ] No (please explain), [**X**] I don't know
Are the results described and interpreted correctly? [ ] Yes, [**X]** No (**see review as text**), [ ] I don't know

Discussion
Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument? [ ] Yes, [**X**] No (**see review as text**), [ ] I don't know
Are the conclusions adequately supported by the results (without overstating the implications of the findings)? [ ] Yes, [ ] No (please explain), [**X**] I don't know

Review The climatic determinants of flexible reproductive seasonality in Africa-dwelling papionins

By Dezeure and colleagues

The main research question of the study is interesting. The authors tried not only to categorize African papionin species into seasonal and non-seasonal breeders and

determine the degree of reproductive seasonality within and among species but also tried to correlate (temporally) certain reproductive events with climatic and NDVI information. The paper is nicely written and the narrative is easy to follow.
The quality of such analyses, however, depends largely on the availability, quality and quantity of data on reproductive events. Quantity and quality can vary largely, e.g., from one species corresponding data of only one population (groups) from 20 years is available, from another species data from 5 populations (groups) for only one or two years, from a third species just data from captivity. Therefore, the question arises how to weigh these differences, or should one ignore them at all? Given this problem, the authors might discuss how such data heterogeneity might affect their results.

Thank you for this relevant comment. Indeed, the birth seasonality quality of the data varies a lot between study sites. We are well aware of this, and that's why we show all sample sizes (the number of births *N* and the number of years of surveys) in Table S1. We agree that such variation can affect the results: the estimation of r_{birth} becomes more robust as more births become available (meaning that adding 1 or 2 births in a large dataset (high N) would not change much r_{birth}, while it might for datasets with fewer births (low N). We explain this in the Discussion, Line 610-613 when talking about the potential limits of our analyses.

Lines 610-613: *'In addition, this study is based on datasets with heterogenous resolutions, including diverse numbers of births and years of study (e.g. Amboseli: N=496, Nyears=33, versus Queen Elizabeth: N=35, Nyears=2). Additional birth records in small datasets may change r_{birth}, and could thus alter some of the results in our study'*

In addition, in order to tackle this problem at best in our analytical approach, we now weigh each data point depending on the sample size (N, number of births), using the 'weight' parameter of the 'brm' function. More precisely, we used standardized, log-transformed values of the sample size N to weigh data points. We performed the log-transformation as it is likely that there are diminishing returns to increasing the sample size (above a certain threshold, increasing N does not affect rbirth as strongly as at small sample size). In our standardization we divided by the smallest value, such that the weight for the population with the lowest number of births was set to 1, whereas the other populations count more in the model (with a maximum weight value of 1.86 for the population with the maximum number of births recorded). See Lines 352-359:

'Given the high variation in sample size (i.e., number of births recorded) between populations, we used a weighed regression, where the weight given to each data point equals to log(N) / minimum(log(N)) so that the population with the lowest sample size counts for 1 observation, and the other populations count for more observations depending on their sample size, following a logarithmic scale ; the logarithmic scale was chosen to account for the diminishing return of increasing the sample size of samples that are already large. Beyond a given sample size, further increases in sample size do not affect much r_{birth} estimates, which are already stable and precise.'

I also think that another general question is when to speak of seasonality. How is seasonality defined? This is a difficult question and the authors used here a circadian statistic parameter (rbirth) to differentiate seasonal from non-seasonal breeders. However, they did not use rbirth directly, but the associated p-value. P 14 L 334-336 “Populations were categorized as seasonal breeders when P-values associated with the Rayleigh test were <0.05, meaning that the null hypothesis of a uniform birth distribution could be rejected”. It might be statistically correct, but since the P-value not only depends on the effect size (here rbirth) but also the sample size, using only the P-values to categorize seasonality is in my view inappropriate. For instance, the distribution of 32 births in Gashaka-Gumti reveals a rbirth of 0.2167 and a P-value of 0.2237 and is thus per the definition of the authors non-seasonal. However, the distribution of 118 births in Gilgil reveals an even smaller rbirth of 0.1873 but a P-value of 0.0159 and makes the population a seasonal breeder. In addition, when comparing the distributions of births over the 12 months of the year for Gilgil and Gashaka, as depicted in Fig 2, it remains rather questionable to categorize the Gilgil population as seasonal and the Gashaka population as non-seasonal. The question remains, how large should the rbirth (effect size) be to qualify for seasonality? Another question arises when the authors speak about evolution, in particular the evolution of a flexible reproductive phenology.

Thank you for this interesting comment, which we considered carefully. We agree that the intensity of reproductive seasonality is a continuum thus it is not ideal to break it into categories. However, from a theoretical perspective, it makes no sense to test if the birth peak is synchronized with the annual food peak in non-seasonally breeding populations (i.e., where there is no clear birth peak). Indeed, including ‘false’ birth peaks that may represent artefacts of small sample sizes would create noise, and make results hard or impossible to interpret. We thus prefer our test to be conservative and include only real birth peaks – in other words, we think it’s important to prioritize specificity over sensitivity when testing H2.

As an example, let’s consider the two populations you mention. Given the large sample size for the Gilgil (or Amboseli) population, we can consider than the birth peak there, although small, is genuine, thus meaningful to include. Conversely, given the low sample size, the ‘birth peak’ of some populations like Gashaka could be artefactual thus not ‘meaningful’. Also, Fig2 shows the proportion of births on the y axis, so that a moderate birth peak appears more intense for populations with small samples size (Gashaka) than for those with larger samples (Gilgil). The statistical limitations you’re highlighting (that small effects are only detected with large samples, and that large effects can go undetected with small samples) are very general in statistics. But overall, we simply cannot think of a more rigorous approach than basing our decision on a statistical test (with the usual limitations of statistical tests), and hope this line of reasoning makes sense.

Based on your comment, we changed the phrasing in the main text (see Lines 370-380) to better highlight that our approach does not really aim to separate seasonal versus non-seasonal populations, but rather to exclude populations with a non-robust birth peak based on the results of a Rayleigh test.

Lines 370 - 380: *'We tested H2 only for those populations for which a significant birth peak can be detected, as it does not make sense to test which period of the reproductive cycle is matched with the annual food peak if there is not a clear seasonal pattern of births in one population. We therefore assessed whether each population had a significant birth peak using the Rayleigh test for circular statistics, more precisely the 'r.test' function from 'CircStats' package (Agostinelli & Lund, 2018). For each population, when the P-value associated with the Rayleigh test was <0.05, meaning that the null hypothesis of a uniform birth distribution could be rejected, the birth peak was considered significant. With this approach, some populations with relatively low reproductive seasonality (low r_{birth}) but with a large number of births are included, which should be taken into account when interpreting the results. Among all populations with a significant birth peak (see Table S1), we investigated which reproductive stage (H2.1, 'conception'; H2.2, 'lactation'; or H2.3, 'weaning') was synchronized with the annual NDVI peak, i.e. μ_{NDVI}.'*

P4 L90-91 "flexible reproduction may be more advantageous than strictly seasonal reproduction"
What do you mean by flexibility? On an individual basis or population basis. It might be difficult for a female to decide whether to mate or not if she then has no clue what the conditions will be in 6 months.
P18 L422-424 "Such flexibility would imply that each population, or even each individual may be able to adjust phenology to current environmental conditions."
Should this mean, that an individual female may decide that this year I will start later with reproduction because it seems that the year will be difficult? Usually, such "flexibility" is found at the population level. Some individuals will reproduce at that time of the year and others at another time. If there is selection pressure on the timing, those individuals who reproduce at the optimal time would have a fitness benefit. We can find flexibility within and among individuals, and within and among populations. It would be very helpful here, to get some more information on how the authors think such a process would work.

Thank you for this interesting point. We see flexibility mostly as an individual trait. In the chacma baboon population that we study in Namibia, at Tsaobis, our previous work shows that there is flexibility in birth timings at the individual level. Flexibility at the individual level means that a same female can give birth at different timings for successive birth events, depending on her own individual traits or physiological constraints, or alternatively depending on the strategies of other females in the same social group (see Dezeure et al. 2021, PRSB; Dezeure et al. 2022, Am Nat). This reproductive flexibility at the individual level shapes patterns of reproductive seasonality at the population level, leading to low seasonality. Reproductive flexibility at the population level could be defined as the ability for different populations of a same species to exhibit different patterns of reproductive seasonality. Our analysis suggests that such population-level flexibility occurs in *Papio ursinus*, where populations vary in the intensity of reproductive seasonality (rbirth=[0.10-0.41]), and to some extent, in the timing of the birth peaks (Moremi: around November, versus Drakensberg: around September). However, such flexibility at the population level is usually difficult to assess because often data are available for only one population per species.

We integrated this comment in Discussion, Lines 520-541:

'Importantly, the key adaptive trait that evolved in the Papio *genus may not simply be the loss of breeding seasonality per se, but the evolution of a flexible reproductive phenology. A same papionin female can give birth at different timings for successive birth events, depending on her own individual traits or physiological constraints, or alternatively depending on the strategies of other females in the same social group, as shown in Papio ursinus (Dezeure et al. 2022, Dezeure et al, 2023). This reproductive flexibility at the individual level necessarily shapes population patterns of reproductive seasonality, leading to lower reproductive seasonality. Reproductive flexibility at the population level could be defined as the ability for different populations of a same species to exhibit diverse patterns of reproductive seasonality, depending on the environmental conditions. Such flexibility is observed at the population level in Papio ursinus, living in a large distributional range characterized by exceptional ecological diversity, which includes cold and temperate climates, oceanic and mountainous ecosystems, and tropical and arid savannahs. Indeed, populations of this species exhibit a wide range of intensity of reproductive seasonality (r_{birth}=[0.10-0.41]), with significant (Moremi) or non-significant (Tsaobis) birth peaks, and to some extent, various timings in their birth peaks (Moremi: around November, versus Drakensberg: around September). However, such population-level flexibility is often difficult to assess in many species, given the datasets available (Papio ursinus is indeed the only species in our sample represented by more than three populations with a reasonable number of births).'*

Note that we also defined reproductive flexibility the first time it appears in the main text, in the Introduction section, Lines 93-95: *'In locations with intense year-to-year environmental variation, a flexible reproductive phenology (i.e., individual ability to start a reproductive cycle at different timings of the year, in response to internal or external factors) may be more advantageous than a strictly seasonal reproduction (Brockman & van Schaik, 2005a; van Schaik & van Noordwijk, 1985).'*

Aim II of the study is to tackle the question: What are the main environmental factors responsible for variation in the intensity of reproductive seasonality?
This suggests that the authors try to look for causal relationships between environmental factors and reproductive seasonality. I think this is not possible at this stage. It is more of a correlative relationship that the authors describe.

Thank you for this comment. We agree that our analytical approach, which is clearly correlative, cannot identify a causal relationship. We have now been more cautious in our phrasing, and have changed the title accordingly. We also discuss the fact that a slow life history, which could be selected by environmental unpredictability, could also be an important driver of flexible reproductive seasonality. See Lines 582-584: *'As such, it is likely that rainfall unpredictability selected these traits in papionin species (energy storage, omnivorous diet, slow life histories, etc.), which in turn contributed to shape their flexible reproductive seasonality.'*

For the title of the article, we have now proposed: "Flexible reproductive seasonality in Africa-dwelling papionins is associated with low environmental productivity and high climatic unpredictability"

In my view, the paper has good potential. If the authors will be able to be a bit more critical with their study (data quality, definition of terms, definition of seasonality etc) the study might become a valuable contribution to our knowledge of the seasonality of primate reproduction and will further reveal that most of the African papionins are relatively independent of seasons when reproducing.

Minor comments
P6 L138 primate taxon instead of primates

Done

P7 L169 Where do the data for Rungwecebus come from? I doubt there are data available, not even from captivity.

That's right. We tried to look for data on *Rungwecebus* but did not find any as it is not studied (See Lines 181-185 and Fig.1). We also added the fact that empty circles in Fig1 mean that we did not find any reproductive seasonality data in this species (see Line 192).

'Species for which we could not find birth seasonality data are represented with smaller icons and names, as well as empty circles.'

P8 L181 A definition or description of what the authors mean by "birth peak" and "intensity of birth seasonality" would be helpful.

Yes, we agree. The use of the term 'birth peak' in this sentence was confusing, and we have now rephrased. By birth peak, we meant the mean birth date of a population, i.e. when births mainly occur during the year in the population. By intensity of birth seasonality, we meant to characterise the strength of the seasonality in births, i.e. how seasonal are the births in one population.

We clarify it Lines 202-205: *'We were interested in quantifying two components of reproductive seasonality: the mean population birth date, i.e., describing when most births occur during the year, and the intensity of population birth seasonality, i.e. describing how seasonal the births are.'*

In addition, we now defined 'birth peak' in the Introduction, Lines 65-68:

'These studies have typically detected a relationship between geographic latitude and birth seasonality, suggesting that at higher latitudes, birth seasonality is more pronounced, with a more intense birth peak (a birth peak being the temporal period in the annual cycle during which most birth occur).'

P14 L345 "females tended to conceive during, soon before, or soon after the annual food peak" NDVI is a relatively good measure of plant productivity, however, whether this always coincides with the annual food peak is at least debatable. In temperate zones, NDVIs might be highest at the beginning of summer, but many trees carry fruit later in the year. The more or less strong relationship between NDVI and food peaks (phenology data) needs at least to be addressed.

Thank you, yes NDVI is a proxy of plant productivity that may be, for some species or population, a poor proxy of food availability, as added in the Method (Lines 254-256) and Discussion (Lines 646-650).

Lines 252-254: *'Nevertheless, the use of NDVI as an index of food productivity in papinions, which are not herbivorous (except for Theropithecus gelada), could be arguable, and results of this analysis would be discussed accordingly.'*

Lines 646-650: *'First, although NDVI is a relatively good measure of plant productivity, highest values of NDVI do not necessarily coincide with the annual food peak, especially when focusing on omnivorous/frugivorous species. Precise phenological data from each population would be more accurate to quantify the annual food peak.'*

P15 L358-359 "… closely related species have more similar patterns of reproductive seasonality. However, such a value is also compatible with some phylogenetic flexibility." What do you mean here? If phylogeny has a strong impact, can we then speak of the adaptation of species to certain ecological conditions?

Thank you for this constructive comment, which forced us to reflect on the value and meaning of the phylogenetic signal of reproductive seasonality. We changed the way to compute the phylogenetic signal (see in the method, lines 364-367), as our previous measure 'lambda' is only fully interpretable for traits that follow a simple model of evolution (see discussion in Blomberg et al. 2007 10.1111/j.0014-3820.2003.tb00285.x). Therefore, we now compute the metric of Blomberg's K, which facilitates the comparison of the phylogenetic signal for reproductive seasonality with other traits, such as behavioural or life-history traits.

Lines 364-367: *'We extracted the phylogenetic signal in our dataset with the metric of Blomberg's K, allowing us to compare it with other signals from other traits. To do so, we computed the mean r_{birth} per species, and use the 'phylosig' function with 1000 simulations from the 'phytools' package (Revell 2024).'*

We find a phylogenetic signal K of 1.83. It is relatively high, among the higher values found in the study by Blomberg et al. (see Fig. 6), that examines phylogenetic signals in a range of behavioural, physiological and life history traits in a wide range of organisms. While values of r_{birth} are more similar among species that are more closely related, there is still variation between closely related species. The high signal reflects that reproductive seasonality is shaped by many different factors that are shared among species over evolutionary times. This can also reflect that closely related species in our sample experience the same ecological conditions, so they all adapted to these conditions in a similar way.

We added this interesting point in the Discussion, Lines 510-519: *'The estimated phylogenetic signal is significant, which shows that the intensity of birth seasonality is more similar among two closely related species. The value of this phylogenetic signal (Blomberg's*

K=1.83) is relatively high, ranging among the highest of many life history traits (such as age at maturity, adult mortality, clutch size in birds, sexual dimorphism, etc.), and higher than behavioural traits (such as daily movement distance, prey size, preferred body temperature, etc.), that are more labile (Blomberg, et al. 2003). Despite this strong impact of phylogeny on the intensity of reproductive seasonality, our study emphasizes variations in birth seasonality among closely-related species, or even within a single species.'