



Dr Lewis G. Spurgin  
BBSRC Research Fellow  
School of Biological Sciences  
University of East Anglia  
Norwich  
Norfolk, NR4 7TJ  
United Kingdom  
  
L.Spurgin@uea.ac.uk  
lewisspurgin.wordpress.com

**July 2017**

Dear Ben,

Thank you for taking the time once again to read our manuscript and response letter. Once again the reviews have been robust but fair, and we have been very impressed at the quality of the reviewing process. We have now been through the second round of Associate Editor and Reviewer comments, and adjusted our manuscript accordingly. In particular, we have added further caveats to our discussion with regards to repeatability, and we agree that this does not take away from the positive aspects of the manuscript.

In addition to the revised manuscript, below we provide a point-by-point response to the Associate Editor and Reviewer comments. As with our previous submission, the editor/reviewer comments are in bold, and edited manuscript sections are reproduced in italics and indented.

Please let us know if you need anything else.

Kind Regards,

Lewis Spurgin (on behalf of all authors)

## Associate editor comments

---

One of the original reviewers and myself have now read your revised manuscript and we both think that it has improved substantially. The reviewer, however, very nicely describes some points that I strongly agree would need more attention in the discussion of the paper to more honestly balance/describe the strong AND weak points of your study. These weaker points do not decrease the impact, novelty or relevance of your study, but discussing them in more detail could rather inspire advances to the methodology used in your field of research.

RESPONSE: We are glad that you both agree that our manuscript has improved, and thank you again for your constructive comments. We agree with the central point made by reviewer and yourself, and have now adjusted our discussion accordingly. We hope you agree that this has further improved our manuscript, and we are hopeful that this improved version will generate discussion about both the ecological and technical aspects of telomere dynamics in wild populations.

## Reviewer 1

---

Line 441: It is very kind of the authors to acknowledge me by name, but my first name is Antoine, not Eric. . .

RESPONSE: We are very sorry about this, and we do of course know that your name is Antoine! We have now changed the acknowledgements.

---

As far as I understand ICC, it is relatively close to a pearson r with some adjustments. Therefore, to quantify the proportion % of variance explained, you should square transform it to obtain something looking like a R2. Therefore, the repeatability of 0.68 mentioned by the authors would mean that 46.2% of their measurement is explained by the sample and that 53.8% is explained by measurement error.

RESPONSE: We feel that we will have to agree to disagree here. ICC is calculated as:

$$\text{variationbetweenegroups}/(\text{variationbetweenegroups} + \text{variationwithingroups})$$

Thus if within-group variation is extremely small (i.e. no measurement error), the ICC would be 1, so 0% of the variation would be explained by measurement error. We would therefore interpret an ICC of 0.68 as 68% of the variation being explained by sample identity and 32% by variation other than sample identity (i.e. measurement error).

However, this is a debate over semantics. In the manuscript, we are careful to phrase things in terms of repeatability and avoid using pearson R and R2 values as these are generally considered inappropriate for measuring repeatability. Nowhere do we state that 68% of the variance can be explained by sample identity.

---

While I understand that measurement error is higher in their study due to the long-term nature of the data, such measurement error (if I am calculating/understanding the numbers properly) is still a major issue to underline with more emphasis in the paper. I agree that it will not lead to false positive, but still it is reducing the power to detect any relationship with other factors. In Bebbington 2016, the repeatability given is 0.94 despite having 1064 samples from the same population/dataset. What is explaining such a loss of repeatability between the two datasets?

qPCR efficiency: An efficiency of 78% means that the qPCR conditions have not been optimized properly. Correcting with LinReg for differences in efficiency is minimizing noise related to such problem, but not fixing the problem when efficiency is so different from 100%. Again, this is likely to increase measurement error, but not to create false positive. Such limitation should be discussed in the paper too. The fact that others are publishing with poor efficiency is not a good answer in my opinion, and I can cite you many papers reporting efficiencies being in the range 90-110%, including most (all?) papers of Criscuolo's lab and Monaghan's lab. In LinReg, are you correcting for sample specific efficiency or per the average plate efficiency? Correcting for sample specific efficiency has been criticized in the past : "Sample specific PCR efficiency estimation has its usefulness, but currently only for outlier detection", Hellemans et al. 2007 qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data.

RESPONSE: We are glad that the reviewer agrees that our observed repeatability/efficiency values will not lead to false positives. Our dataset includes repeats that have run several years apart, and these repeats are included in our repeatability estimates. We suspect that our dataset is fairly unique in this respect at present, but other studies will certainly have such data available in the near future. Thew Bebbington et al. paper only included the set of most recently run telomere samples, hence the higher repeatability. We agree that this issue deserves more discussion space, and have now devoted an entire paragraph to the issue of repeatability, and the specific issue of qPCR efficiency, in the discussion on lines 414-428:

*It is worth considering the fact that the social and ecological variables we tested here explained only a small proportion of the variance in RTL. Furthermore, some factors such as territory quality and social group size were not related to telomere dynamics when we may have expected them to be (Van de Crommenacker et al. 2011; Brouwer, Richardson, and Komdeur 2012). Measurement error is clearly an issue in our study, and has almost certainly decreased the explanatory power of our models, and elevated levels of Type II error. The low repeatability we observed is the product of i) lower levels of efficiency in the telomere qPCR reaction than we would have liked, ii) the long-term nature of the study. Samples for this study were run over a period of several years, during which reagents, consumables and personnel all change. Our repeatability estimate includes samples that were run several years apart, and reflect all the sources of error that accumulated over that time. Compared to a set-up where a small amount of samples are all run at the same time, it is unsurprising that we have a higher error rate. Techniques for measuring telomere length with a greater degree of precision are likely to prove helpful in future ecological studies of telomere dynamics (Nussey et al. 2014), and discussions are now required on how to best optimise measuring telomere length for long-term studies. A central issue to resolve is how best to balance the trade-off between obtaining precise telomere measurements, and utilising the large sample sizes necessary for ecological study.*

---

**Information on the age of birds:** While I acknowledge that the authors now provide better information on the age of individuals, the resolution they have for TL at the nestling age is not optimal in my opinion since many research groups (Criscuolo, Monaghan, Verhulst) have identified telomere shortening within 1-4 weeks in nestlings, and the data the authors present here seems to indicate that most TL shortening occur early (very early?) in life.

RESPONSE: We would love to have had fine-scale longitudinal samples from within the nest, but this is not possible in the Seychelles warbler system, in which the nests are very high and difficult to access safely. Systems in which birds can be sampled repeatedly early in life (e.g. from nest boxes) are much better for addressing the specific issue of early life telomere loss. The point that we tried to make was that the (lack of) sampling resolution in early life should be offset against having broad-scale, longitudinal samples from a large number of cohorts. Nonetheless, we agree that our set-up is not optimal for looking at early-life telomere shortening, and have added a section to the discussion on this, lines 328-337:

*Our study adds to the substantial body of literature from humans and wild animals showing that telomere length decreases with age, and that this decrease is most rapid in early life. Rapid telomere shortening in early life occurs as a consequence of the much more rapid rate of cellular division that occurs during the growth phase, but also perhaps higher levels of cellular stress during development (e.g. Frenck, Blackburn, and Shannon 1998; Haussmann, Vleck, and Nisbet 2003; Heidinger et al. 2012). Because we are rarely able to sample Seychelles warblers more than once in the nest, our system is not the best suited for looking in detail at the reasons behind telomere shortening during the growth phase. In contrast, we have good longitudinal and cross-sectional sampling from across individual lifespans, and using this data*

*we have shown that, despite an overall trend for shortening, telomere length both increased and decreased, especially after the juvenile period.*

---

Intra-individual repeatability of 8%: Maybe “worrying” was not the proper word; please excuse my approximate usage sometimes of English. Yet, I am still ‘puzzled’ by the very low intra-individual repeatability for telomere found here and by Dan Nussey’s team. I don’t have similar long-term dataset, but overall in my datasets, I find an intra-individual repeatability ca. 65% over a period of 9 days during growth in tits (see Fig below) for great and coal tits from Stier et al. 2016 J Avian Biol), and 28% over a period of 105 days in king penguin chicks (unpublished). Verhulst’s team find similar/stronger intra-individual repeatability (e.g. Boonekamp 2014 both for nestling and between nestling and adulthood; Bauch 2013 Mol Ecol in adult seabirds).

Do you have any idea about the presence of interstitial telomeric sequence in your study species? Because presence of ITS would even artificially increase intra-individual repeatability when telomeres are measured with qPCR. If there is only 8% of intra-individual repeatability, it is quite surprising that telomere length at a given point can predict survival, as shown by many studies including in your own study system. With such a low intra-individual repeatability, I agree that telomeres can be seen as a biomarker of short-term “costs”, but we would have to re-think our way of using it as a biomarker of long-term costs and survival prospects. This idea should be better highlighted in the discussion in my opinion.

RESPONSE: We thank the reviewer for clarifying this point. Unfortunately we do not have any data on the presence of ITS. However, we have now added further discussion on what may cause the difference between our study and the ones mentioned above. We have also now discussed what low intra-individual repeatability may mean for the relationship between telomere length and long-term survival - lines 364-380:

*The finding that increases in telomere length may be sporadic and overlaid on an overall pattern of shortening with age is an important point when assessing the occurrence of telomere lengthening. Previously described approaches to distinguish telomere elongation from measurement error, based on assumptions about follow-up time between measurements (Steenstrup et al. 2013), or based on measuring variance among measurements (Simons, Stulp, and Nakagawa 2014), assume that telomere elongation within individuals is consistent over time. Our data, and that of others (Fairlie et al. 2016; Hoelzl, Cornils, et al. 2016; Hoelzl, Smith, et al. 2016) suggest that this is not the case. Such inconsistent changes in telomere length over lifespans could occur due to changes in the cellular composition of the blood within individual samples, or due to the actual elongation of telomeres (Blackburn et al. 1989). Determining the mechanism of these changes is essential for how we view telomeres as biomarkers of costs. For example, if telomeres can be lengthened in response to improvements in environmental conditions, this would suggest that they reflect short- to medium-term costs, rather than the cumulative costs that an individual has faced over its lifespan (Bateson 2016). It is clear that telomeres can be a marker of long-term as well as short-term costs, as telomere length has been associated with both survival and lifespan in wild populations (Barrett et al. 2013; Stier et al. 2015), but we do not yet know how biologically meaningful within-individual fluctuations in telomere length are. New research is therefore required to determine when and why telomere length increases within individuals, so that biologically informed hypotheses about the nature of telomeres as biomarkers in wild populations.*