The replication of empirical findings is fundamental to science yet it is generally ignored in the biosciences. Over the past decade, however, a replication crisis has prompted large-scale, systematic efforts to replicate foundational studies in psychology and biomedicine. Surprisingly, many of these replications have failed to repeat the findings of previously published influential, high-impact work. Little is known about the extent of these problems in ecology and evolution but there is no a priori reason to think that these fields are immune to the ails affecting other biosciences and psychology. Behavioural ecologists, for example, have a rather poor track record of replicating studies although considerable conceptual replication takes place (i.e. testing the same hypothesis but using different species and methods).

Our aim here is two-fold. First, we attempt to determine how frequently studies are replicated in ecology and evolution by text-mining papers that were published in ecology and evolution journals (more generally Plant and Animal science). Second, we determine whether studies in ecology and evolution claiming to replicate previously published findings indeed do so.

We downloaded as .xml files 1 641 366 Open Access papers available in the PubMed database representing 7439 journals. We used code written in the Python language to text-mine these papers for any permutation of the word “replicate” (i.e. “replic\*”) in the Introduction or Discussion. For each instance of “replic\*” we extracted the sentence as well as the paper’s meta-data (doi, ISSN, etc.). Each of these instances was added as a row to a .csv file.

Results and Discussion

We extracted 16420 sentences containing replic\* from 4462 papers comprising 72 “Ecology, Evolution, Behavior and Systematics” journals (SCimajor 2017). We focused only on papers that tested the relationship between variables or tested the effect of a factor(s) on some response variable. We therefore eliminated all papers (n=474) published in PLoS Computational Biology because these studies did not meet our criteria of repeating a previously published empirical study from which effect sizes can be calculated. Inspection of the sentences from the remaining n= 3988 papers revealed that the majority of papers mentioned “replication” within the context of experimental design (i.e. sampling units) or genetic mechanisms (e.g. DNA replication). A handful of papers (n=25) dealt with study replication in the sense of commenting on the practice of study replication, mentioning replications conducted by others, or suggesting that the current research study be repeated. The authors of one study (Troncoso-Palacios 2016) stated that they could not assess the replicability of their study because the original authors did not provide sufficient data for assessment whereas the authors of another study (Troncoso-Palacios 2015) claimed to have too small a sample size to assess whether they replicated the findings of an original study (Medina et al. 2013). One paper (Schutte et al.) conducted a study and then compared their results to previously published findings using meta-analystic techniques and Amos re-analyzed data previously published by Møller and Cuervo (2003). ~~Only 10 papers (0.25% of all papers) dealt with authors repeating a previously published study~~ ~~but we were able to calculate effect sizes for the original and replicated study in only n=5 of these cases.~~

~~Of these 36 papers, 40 (18 plant/animal and 22 enviro/eco) discussed the general importance of study replication or suggested that the current findings be replicated by others (REFS).~~

Only 10 publications (0.25% of all Ecology, Evolution, Behavior and Systematics papers) claimed to have attempted a replication of a previously published study (REF; Table X). For each of these replication studies we retrieved the original study and extracted the data from both in order to calculate an effect size (Hedges d or Pearson’s r) and 95% CI using conventional methods (Lajeunesse chapter). If the information required to calculate an effect size was not available in the text or tables we extracted data from figures using GraphClick. Some studies replicated several tests in the original study and so we calculated an effect size for each of these. We deemed the replication successful if the 95% CI overlapped the 95% CI of the original study.

Of the 10 papers claiming to have replicated a previously published study n= 5 sets of studies were excluded because either the original study (Jacobs e tal. 2014; Range et al 2011) or neither the original (Jones et al. 2012; Lassance et al. 2010; Nagawasa et al) nor the replication (Mazzarella et al. 2016; Coates et al. 2013; McGreevy et al) provided the data necessary for effect size calculation. Ramirez et al. (2014)/Semm and Demaine (1986).

This means that only five sets studies could be compared for replication success.

Authors of replications assessed the success of their replication based on the direction of the effect and its associated p-value. This method proved surprisingly accurate as five of our five replication comparisons supported the replicating author’s conclusion. For example, Pasukonis; Mendes; Karin-D’Arcy stated that they replicated the original study and indeed they did. Riemer; Bulla et al. (2014) and Fisher and Adams (1981) stated that they failed to replicate the original finding; our analyses support the conclusions of Riemer and Fischer and Adams; however, Bulla et al. 2014, appear to have successfully replicated Cresswell et al. (2003).

Our conservative analysis of the open access literature suggests that approximately 0.25% studies in ecology and evolutionary biology attempt to replicate a previously published study. This should probably be increased. What is the rate in other disciplines?

Is PubMed OA a good guide? Why would replications be less OA than not. I would predict these types of studies would be published at a higher rate in open access journals like PloS. Why would we miss replications – unlikely for someone to conduct a replication and not use the term “replic\*” at least once in the paper.

It’s clear that authors are aware of the importance of replication but unfortunately this aspect is discussed at more than twice the rate (25/3988) than the actual completion (10/3988) of replications.

Junk…

Replicate in the sense of duplicating a phenomenon eg. Mesocosms replicating/mimicking what happens in nature

For example van wijk et al. (2008) discuss replication with respect to experimental design.

Ramirez et al’s (2014) replication of Semm and Demaine’s (1986) study (cited > 150 times) examining magento sensitive cells in the pigeon’s visual system was not included in the forest plot. The original (~79%) and replicated (~0%) study both counted the number of magneto-sensitive cells in the pigeons system and will not yield effect sizes. Ramirez failed to replicate Semm and Demaine.

Independently replicated in a separate study (e.g. Simon 2013 did replications in different centers within the same study).

Trnka and Grim (2013) and Mychaleckyj et al. (2017) and Kalinkat et al. (2013) offer familiar platitudes that things need to be replicated. Replication “qualitatively” affirmed as in Kothapalli et al. (2016).

Goto et al. (2011) explicitly outline how to best repeat their procedure.

NOTES:

Amos was unable to replicate Møller’s statistical analysis despite having the original dataset.

Conceptual replication….

phenomenon Burks et al. 2016 stated that they replicated Steiner et al. 2010 but it was a conceptual replication wherein the function of similar antennal morphologies was studied in two different hymenopteran species. Similarly, Marshall and Stevens (2014) and Kandul et al (2006) performed conceptual replications and stated that their finding replicate findings in other species.

only 5 provided the data or statistics necessary to calculate an effect size. Muller et al. 2014 claimed to have replicated a cognition study in dogs reported by Range et al. (2011) but the latter provided only a Wilcoxon T+ statistic. Troncoso-Palacios 2016 mentioned that relevant data (molecular sequences, morphological characters) were not provided in original studies (e.g. Pincheira-Donoso and Núñez, 2005; Avila et al. 2010) and so the study (phylogeny building) could not be replicated. Troncos-Palacious et al. 2015 state that they could not replicate the analyses of Medina et al. 2013 due to small sample sizes.

Finally, Nagasawa et al. 2016 report partial, not raw, correlation coefficients; therefore, we could not determine whether they replicated McGreevy et al. 2013.