Path forward to making stronger tests of mismatch hypothesis- related to the quantitative synthesis of studies

1. Gold standard: *Integrate long-term observational studies with experiments*.
   1. Long-term observational study (i.e. one that starts before early 1980s) that (1) shows clear evidence of baseline conditions (aka stationarity); and (2) has independent and equivalent phenological and fitness measurements for both resource and consumer. Where such data have not been collected before the early 1980s, there is no perfect solution to this challenge. However, null modeling may be an alternative.
   2. Use experiment or some other method (process model) to understand phenological cues (need more details) and limits to fitness both in terms of peak and zero. Experiments are needed to manipulate the timing of the interaction. Gold standard experiments are those conducted against a meaningful baseline and push the system to the limits at both extremes.
   3. Ideally, phenology would not be measured based on population-level metrics. For example, the peak date of activity is based on 50th percentile abundance. Here, the relationship between phenology and fitness is difficult to separate. In aquatic systems, the phenological measurements of the resource would not be impacted by the consumer at the time of measurement.
2. Silver standard: *Comprehensive observational studies or experiments.*

In some systems, it may not be possible to integrate long-term observational studies and/or to do experiments. In these cases (and in the gold standard studies), scientists should strive to have:

1. Higher resolution phenological data. In aquatic systems, there is quick turnaround between producers and consumers- temporal sequencing is difficult to determine therefore sampling frequency is important; AND- some zooplankton not in complete dormancy, some remain at low densities.
2. Full extent of resource variation (e.g. Deacy et al. 2017). Clarity on diet breadth would help calculate interaction strength and thus determine whether one of the main assumptions of the Cushing hypothesis is supported. Interaction strength could influence the degree to which the relative timing of an interaction responds to climate change and the magnitude of those consequences for consumer fitness. In a recent meta-analysis (Kharouba et al.) evaluating the climate change driven shifts in phenological synchrony, only rarely was enough detail provided in the paper to able to quantify the strength of the interaction. and therefore determine whether Cushing hypothesis is even expected in the system.
3. Bronze standard: *Detailed descriptions of system and being explicit about assumptions related to system and Cushing hypothesis.*

In systems when unable to meet the gold and silver standards outlined above, or in addition to those approaches, studies should strive to:

1. When the baseline in the system is not, or is unlikely to be (if unknown), necessarily the same as the peak of the Cushing curve (i.e., peak fitness) (Figure 2).
2. Be explicit about the *direction* of fitness consequences in response to the *magnitude* of change in relative timing. For example, does more time between the life history events of the resource and consumer consistently lead to declines in fitness? This can be challenging in the lower trophic levels of aquatic systems where is there is quick turnaround between producers and consumers (see silver standard).
3. Discuss the likelihood that the major driver of change in phenology in the system is climate change. Authors should be explicit about the role of temperature as a phenological cue for either species.
4. Address the key assumption that the resource is (or is likely to be) the major control on the consumer.
5. Be specific about how phenological phase relates to the Cushing curve. Ideally, the most energetically expensive part of the consumer’s phenology would be measured (e.g.) but when not known or not specified, be explicit that a phenological phase proxy is used instead of actual key event (e.g. use lay date for chick activity) and include an estimate of the time lag between the proxy and phase of interest.
6. Discuss the relative importance of the measure of fitness used in the life history of the species. For example, if adult body size is measured but most mortality occurs in the juvenile stage.
7. Be explicit about which mechanism is likely driving the curve. For example, if individual fitness is the focus, then the mechanism will be related to life history theory and could be X,Y, or Z.
8. Collect data on fitness and equivalent data on the consumer and resource.