**Title:** Trophic phenological mismatch: Disconnects between underlying ecological theory and climate change responses

**Authors:** Heather M. Kharouba1, Elizabeth M. Wolkovich2,3

**Affiliations:**

1Department of Biology, University of Ottawa, Canada K1N 9B4

2Harvard

3UBC

Email: heather.kharouba@uottawa.ca,

Running Title:

**Keywords:**

**Type:** Ecology Letters (Ideas & Perspectives)

**Abstract**

Many researchers hypothesize that climate change will lead to phenological mismatches with negative consequences for those interacting species and their ecological communities; yet, evidence documenting negative impacts on fitness is mixed. The most common ecological theory that underlies these studies is the Cushing match-mismatch hypothesis. It offers a testable hypothesis that—when combined with a system’s pre-climate change baseline condition—can predict these consequences due to climate change. However, none of the studies we reviewed collected the data required to provide strong tests of this hypothesis, making it difficult to assess support for this major hypothesis. Further, 74% of studies fail to define pre-climate change baselines in their study system, making predictions about climate change impacts on consumer fitness difficult. To accurately predict the magnitude and prevalence of mismatches due to climate change, relating empirical observations to underlying mechanisms through hypothesis testing will be required. By adjusting their study designs, researchers can more rigorously test this hypothesis. We highlight how these approaches could rapidly advance our mechanistic understanding and thus allow robust predictions of shifts with continuing climate change.

**Introduction**

Climate change is causing phenological shifts (i.e. changes in the timing of life history events) that vary across species in different functional groups and trophic levels (Thackeray et al. 2016; Ovaskainen et al. 2013; CaraDonna et al. 2014). Such species-specific variation in response to climate change has led to changes in the relative timing of key activities (phenological synchrony) among interacting species (Kharouba et al. 2018). These changes have caused fitness consequences—often termed ‘phenological mismatch’ (Box 1)—and have influenced ecosystem-level properties in some contexts (Post and Forchhammer 2008; Plard et al. 2014; Doiron et al. 2015; Burkle et al. 2013), but not others (Vatka et al. 2011; Burthe et al. 2012). Recent theoretical (Johansson et al. 2015; Bewick et al. 2016) and empirical studies (e.g. Samplonius et al. 2016) based in single systems have worked to improve predictions and address diverse findings. Yet—while there is general agreement that predicting phenological mismatches is critical for determining the extent to which pair-wise species interactions, communities, and ecosystem function (e.g. pollination) will be affected by climate change—we still have no general ability to predict the outcomes of shifts in phenological synchrony due to climate change.

Here, we argue that much of the difficulty in predicting the consequences of climate change-driven shifts in synchrony is due to a disconnect between ecological theory and current empirical approaches in the phenological mismatch literature. Current methodological inconsistencies across studies make it difficult to test the relevant underlying ecological theory in the context of climate change. Without better evidence, we cannot attribute variation in findings across studies to species, site, or mechanism. Without an understanding of the mechanisms underlying the well-documented patterns in phenological shifts, our ability to make accurate predictions about species’ responses, and species’ interactions, to climate change remains limited (O’Connor et al. 2012; Chmura et al. 2018).

Here, we focus on the widely-cited Cushing match-mismatch, or trophic mismatch, hypothesis (1974), the most commonly applied hypothesis to consumer-resource interactions in this literature. We show how advances could come from direct tests of the hypothesis and clear definitions of baselines, when possible. Our aim is not to put forward additional hypotheses about the context in which phenological mismatch will occur, which has been reviewed extensively elsewhere (e.g., Miller-Rushing 2010; Renner and Zohner 2018), but rather to help guide the study of phenological mismatch to develop robust predictions that can scale up to inference across sites and systems by reviewing the major competing hypotheses in the literature.

Although the Cushing hypothesis has been applied to other types of interactions (e.g. mutualism), we limit our discussion to consumer-resource interactions (i.e. antagonistic). Below, we provide an overview of the Cushing hypothesis and then outline the divide between the hypothesis and the empirical studies using a systematic literature review of phenological mismatch. We then discuss how current approaches are impeding major progress in the field, but how changes in our approach could rapidly advance our understanding and help forecast of the impacts of climate change on ecological communities—the ultimate goal of most of the phenological mismatch literature.

***Overview of the main ecological theory***

The Cushing match-mismatch hypothesis is the most common ecological theory underling phenological mismatch studies (Appendix). This hypothesis predicts the often-shown concave-down curve between consumer fitness and relative timing between the consumer and its resource (1974; Figure 1). While this curve has been applied across many ecosystems (e.g. Winder and Schindler 2004; Vatka et al. 2011; Arula et al. 2014), the theory originally emerged from the marine fisheries literature as a way to explain the variation in population recruitment of fish stocks.

Based on life-history theory, the Cushing match-mismatch hypothesis postulates that there should be selective pressure for the consumer to temporally ‘match’ the peak of its energetic phase with the peak of resource availability (Figure 1). Given this strong selective pressure, if there is any change to the relative timing of the interaction, there will be a decrease in fitness for the consumer (i.e., a mismatch), thus producing the now-often-shown curve (Figure 1). For example, if a consumer emerges too early in the spring and the resource has not yet emerged, the consumer will risk starvation. At the curve’s limits, consumer fitness should fall to zero when the change in relative timing is sufficiently large (Figure 1). For example, the early-emerging consumer can only survive so many days without food.

This hypothesis is based on two important assumptions. First, the resource is the major controller on the consumer (i.e., strong interaction strength, bottom-up control; Cury et al, 2003). Second, the consumer and resource must present a certain degree of seasonality (e.g., resource is only available for part of the year; Durant et al. 2005; Durant et al. 2007), limiting the optimal period for growth and reproduction for the consumer. If neither of these assumptions is met, then fitness consequences due to changes in the relative timing of the interaction will be either weak or non-existent.

***Disconnect between theory and empirical studies***

In its original state, the hypothesis has been debated, contested and criticized, particularly in the marine literature (Leggett and DeBlois 1994\*; Durant et al. 2007). In part because, although a relatively simple hypothesis, it is inherently difficult to test in the field, an assertion even Cushing himself made. Indeed, the shape and strength of the relationship of the curve varies greatly across observational studies (e.g., Philippart et al. 2013; Reed et al. 2013; Plard et al. 2014; Atkinson et al. 2015).

While others have suggested that this is because of data limitations and the model’s implication of complex multitrophic dynamics (Durant et al. 2007; Kerby et al. 2012), we argue that there are two key theoretical areas that make it difficult to determine whether this hypothesis is widely supported in the context of climate change: studies often fail to rigorously test the Cushing hypothesis and/or test pre-climate change conditions (Figure 2). Below, we introduce the current objectives of the phenological mismatch literature and then discuss these two areas.

To put our argument in context, we systematically reviewed the phenological mismatch literature. From this search, we examined 43 observational studies that met our criteria (see Appendix for details) for evaluating the effects of a change in relative timing on the fitness or performance (hereafter referred to as performance) of the consumer. The majority of the studies (27/43) focused on: i) documenting how climate change is affecting the timing of a trophic interaction and how those changes have affected the consumer’s performance; and ii) advancing the theory necessary to predict long-term demographic changes due to changes in synchrony. The ultimate goal of these studies was to predict the impacts of climate change on ecological communities (hereafter called ‘climate change’ studies). Our database also included studies (16/43) whose aim was to understand the underlying processes related to timing (e.g., coevolution, life history trade offs, food web dynamics) that drive consumer or resource dynamics (hereafter called ‘fundamental studies’), independently of climate change. Combined, these 43 studies have improved our understanding of the importance of the relative timing of an interaction for consumer fitness.

*i) Testing the Cushing hypothesis*

The Cushing hypothesis offers a testable hypothesis for predicting the magnitude and direction of demographic changes in response to climate-change driven shifts in synchrony. To date, much research in the biological impacts of climate change literature has focused on the direct relationships between organisms and the abiotic environment (e.g., Menzel et al. 2006, Chen et al. 2011) rather than testing theory (Lavergne et al. 2010; O’Connor et al. 2012; Mouquet et al. 2015; Barner et al. 2018). However, progress on the Cushing hypothesis—rigorously testing the assumptions and ultimate mechanisms (Figure 2)—requires tests of a diversity of ecological and evolutionary theory. This represents the major challenge of the hypothesis and—we argue—may be why support for it has been so mixed.

Many mechanisms can produce the Cushing curve, and they vary across systems, space and time. Here we focus on the ultimate mechanisms related to life history theory and food web theory. Mechanisms arising from life-history theory (Box 1), for example, trade-offs between fecundity and mortality, and bet-hedging strategies (Figure 2), can produce the Cushing curve—as consumers maximize their fitness through ideal timing with their primary resource—this forms the theoretical basis for the original hypothesis. The Cushing hypothesis, however, is also connected to food web theory which focuses on predator prey dynamics (Box 1) and is related to the timing of a consumer to its food resource. If the most likely mechanism in the system violates one of the key assumptions of the hypothesis, then the Cushing hypothesis is unlikely to apply in this system (Figure 2). For example, if mechanisms, such as density dependence or top-down pressure, are thought to be at play, then the first assumption of the hypothesis is unlikely to be met.

Studies deviate in what forces they hypothesize control the peak in the food resource (a critical component of the Cushing curve and a key assumption, see Figure 1 and 2).

For example, in aquatic systems—where top-down forces are generally more influential compared to terrestrial systems (Shurin et al. 2005)—many studies suggest that the resource peak is controlled by release from, or predation by, a consumer (Carpenter and Kitchell 1996; Shurin and Seabloom 2005; Borer et al. 2006). This is a very different hypothesis from others that suggest seasonality in the environment produces the resource peak (Hampton et al. 2006). In terrestrial systems, the relative contribution of direct vs. indirect effects on consumer dynamics remains poorly understood (e.g. Boggs and Inouye 2012) and these two hypotheses can result in different types of effects on population dynamics.

The diversity of ecological theory that can produce the basic shape of the Cushing curve means that different researchers may collect very different data depending on what mechanism they suspect underlies the curve in their particular system or what their main objective is. Researchers studying the hypothesis from the lens of life history theory require per capita estimates of fitness, including measurements of fecundity and mortality, and how they vary across ontogeny; such measurements allow researchers to assess fitness consequences at the level of selection (i.e., the individual) and to directly link any fitness changes to timing. Studies addressing the Cushing hypothesis from the perspective of food web theory require more equivalent data on the consumer and resource but do not require per capita fitness consequences for the consumer.

Strong tests of the Cushing hypothesis include clear tests of the assumptions alongside tests of multiple mechanisms. To accomplish this, performance metrics would ideally be assessed for both the consumer and resource, and at the level of the individual when possible, but such data are non-existent in this literature (i.e. none of the pair-wise interactions had individual-level data for both the consumer and resource). Of the interactions we examined with per-capita performance data on the consumer (i.e. life history; n=28 interactions), none had equivalent measures of performance on the resource (Table 1). For the majority of these interactions (18/28), the researchers measured performance at the community level for the resource (e.g. biomass; Table 1) and many had no measure at all for the resource (6/28 interactions; Table 1). As expected, researchers that measured population- or community-level data on the consumer (i.e. food-web; n=18 interactions) collected more equivalent data on the resource and consumer. For the majority of these interactions, researchers also collected population-level (7/18) or community level data for the resource (7/18) (Table 1). Across all life history and food-web interactions, researchers measured per-capita performance on the resource for only one interaction (Philippart et al. 2003 (AO001)).

Part of these discrepancies may stem from differing generation times—studies that can easily observe a resource curve shaped by predation often focus on organisms with generation times on the scale of days to week. For example, aquatic studies that focus on phytoplankton as the resource may observe many generations in one summer while a terrestrial study focused on caterpillars would generally observe a comparably smaller number generations (1-2(?), depending on latitude and species for caterpillars). When populations turnover too quickly to track individuals, researchers struggle to collect robust data on per capita fitness. The fundamental difference in the scale of generation times pervades aquatic/terrestrial comparisons, particularly at lower trophic levels (Borer et al. 2005; Shurin et al. 2005; Gruner et al. 2008) and may drive the difference in the objectives of aquatic vs. terrestrial studies in the phenological mismatch literature. Indeed, the majority of researchers studying aquatic-based interactions in our review approached the Cushing hypothesis from the perspective of food-web theory (15/20 interactions), whereas researchers approached terrestrial-based interactions from life-history theory (23/26; Table 2).

A further divide across studies may come from the size, and population size, of the consumer, properties correlated with generation time: it can be difficult to collect per capita fitness data when consumers are too small and numerous. Of the life history interactions, the majority of consumers were birds (19/28) and mammals (6/25; Table 2). Comparatively, invertebrates were measured at the population or community level and the vast majority were aquatic (11/12) (Table 2). Therefore, only certain classes of organisms are tested for each category of mechanisms and in one type of biome, further limiting generalizations across systems.

This places the phenological mismatch studies on a continuum: at one end, aquatic studies focus generally on food web mechanisms, collect equivalent data on the consumer and resource but no per-capita fitness data and at the other end, terrestrial studies focus on life history theory from the lens of the consumer, collect individual-level data on the timing and fitness of the consumer, with much less information on the resource. The fundamental problem with these approaches is that researchers test only one piece of the much larger field of mechanisms that could underlie the Cushing curve. Consequently, most studies do not actually provide strong tests of the Cushing hypothesis, making it difficult to refute the hypothesis if no evidence is found, a prevalent issue in ecology and evolution (Betini et al. 2017). If they find their mechanisms explain little of the variation they observe, they rarely have the data to test alternative hypotheses.

Without strong support for their hypothesis (i.e., high explained variation and clear underlying links), the mechanism underlying the curve will be uncertain. This uncertainty can be caused by two reasons: (1) hypotheses that are not mutually exclusive (i.e. different mechanisms have the same predictions). For example, the relative timing of an interaction will change in the same direction regardless of whether temperature (or some other abiotic factor) directly or indirectly affects a resource’s peak; or (2) considering multiple mechanisms yields different predictions than considering only a single mechanism. For example, researchers have pointed out how a broader perspective on life history, such as including other sources of mortality beyond those related to resource access or alternative resource types (Samplonius et al. 2016), can lead to alternative predictions than the synchrony predicted by the Cushing curve (Singer & Parmesan, Johannsson & Jonzen 2012).

Finally, another consequence of researchers not collecting equivalent fitness and population data for both the consumer and resource is that our mechanistic understanding ends at the consumer level. Without finer scale data on the resource, it will be difficult to test the 2nd assumption of the hypothesis and make predictions of the broader consequences of climate change-driven phenological mismatches for the interaction in its entirety. For example, resource abundance can be used to determine the likelihood that consumers will encounter enough resource at the tails of their temporal distributions, thus defining the degree of seasonality in the resource. Moreover, we will not be able to incorporate feedbacks between the consumer and resource (e.g., Carpenter et al. 2001; Thackeray 2012), measure interaction strength (Miller-Rushing), which informs a key assumption of the hypothesis, or determine the broader implications of phenological mismatches for the community (e.g., Nakazawa and Doi 2012; Revilla et al. 2014).

*ii) Testing pre-climate change conditions*

Applying the Cushing hypothesis to understanding—and ideally predicting—the consequences of climate change for interacting species requires more than testing the assumptions and underlying mechanisms of the hypothesis (Figure 2). If the assumptions of the Cushing hypothesis are met and the mechanisms understood, researchers still need an understanding of the system dynamics before climate change, and the phenological cues for consumer and resource levels, to predict what shifts in the climate have—and will—do to the timing and fitness of the players (Figure 2). Below we discuss these steps in more detail.

*(a) Identifying pre-climate change baselines*

The concept of pre-climate change baseline (i.e., defining the range of natural variation) is rarely, if ever, considered by phenological mismatch studies; yet, establishing one in a system has important implications for the Cushing hypothesis in both climate change and fundamental studies. Of the studies we considered, only 26% (11/43) of the studies began before the early 1980s, the same time that pronounced recent climate change began (Solomon et al. 2007). Even in this subset of studies, not all studies had more than three years of data (2/11 had only one or two years of data for inference). This is also an issue with the phenological synchrony literature (Box 1), which does not necessarily consider the ecological consequences of shifts in synchrony (Kharouba et al. 2018). Even those studies with time-series that extend into decades before the 1980s do not explicitly define a pre-climate change baseline in synchrony (but see Philippart et al. 2003 (AO001); Adrian et al. 2006 (HMK031), Arula et al. 2014 (HMK002)).

There are several major implications of not establishing a pre-climate change baseline (Figure 3). First, studies without a baseline are implicitly assuming conditions of stationarity (i.e., a constant underlying probability distribution) before climate change occurred (Figure 3a; Wolkovich et al. 2014). This is problematic when climate change has led to non-stationarity to different extents in different systems (Figure 3a). Without a defined pre-climate change baseline in the system, we do not know whether, and to what magnitude, we should predict a divergence in synchrony due to climate change (Figure 3a). It is then further difficult to determine how much of the observed change in the interaction can be attributed to climate change, some other driver, for example nitrogen deposition (Edmondson 1994), land use change (Sala et al. 2000), species introductions (e.g., Ricciardi et al. 1998; Fritts and Rodda 1998; Verschuren et al. 2002;), or natural variation. Thus, knowledge of the pre-climate change baseline may rule out climate change and identify other drivers, which is key for the prioritization of conservation management decisions for threatened species.

Second, these studies confound a pre-climate change baseline with the peak of the Cushing curve: that is, they assume that fitness was highest before climate change because the consumer’s timing was well matched to the resource’s (i.e. the synchrony hypothesis; Box 1; Figure 3b; Singer and Parmesan 2010). The resulting prediction is that climate change will necessarily change the relative timing of the interaction (i.e. non-stationarity) leading to a decline in the consumer’s fitness (Figure 2, Figure 3b). However, without knowledge of the pre-climate change baseline, it is difficult to predict the direction and magnitude of fitness consequences due to climate change-driven shifts in synchrony, especially when alternative conditions could be the baseline (Figure 3b,c).

An alternative hypothesis put forward by Singer and Parmesan (2010) suggests that in some contexts, ‘asynchrony’ (i.e. when the most energetically demanding phase of the consumer is not lined up with the peak resource availability; Box 1) may be the pre-climate change baseline (Figure 3b). In systems where asynchrony or other conditions might be the baseline state, our ability to anticipate the effects of climate change will be even more challenging as these alternative baselines could lead to vastly different predictions under climate change (Figure 3c). For example, interactions might become less (i.e. fitness increases) or even more mismatched (i.e. fitness decreases) with climate change (Figure 3bc).

The pre-climate change baseline is dependent on the ultimate mechanism(s) of the Cushing hypothesis (Figure 2). For example, a consumer that maximizes survival over fecundity is predicted to have a pre-climate change baseline of synchrony, whereas a consumer that maximizes fecundity over survival is more likely to have a baseline of asynchrony (Singer and Parmesan 2010). Therefore, researchers need an understanding of the system dynamics before climate change.

Establishing a pre-climate change baseline also has implications for fundamental studies that aim to understand the underlying processes of consumer-resource dynamics. A pre-climate change baseline would determine whether the system is most probably in equilibrium or transient dynamics. For example, life-history trade-offs, including those that can cause the Cushing hypothesis, predict some form of equilibrium conditions, but if climate change has pushed the system away from a previous baseline, then the system might likely be in a transient dynamics phase rather than at equilibrium (Figure 3). To this end, an understanding of the system’s pre-climate change baseline state is important for providing context and even designing fundamental studies.

*(b) Identifying proximate cues*

Predicting how climate change will affect the relative timing of a consumer and fitness, and potentially the fitness of the consumer will also depend on the identification of the proximate phenological cues of the consumer and resource (Figure 2). For example, climate change is likely to lead to fitness declines for a consumer that has a pre-climate change baseline of synchrony and that shares similar environmental cues with its resource (Figure 2). Currently, relatively little is known about the similarity of cues for interacting species across trophic levels (Chmura et al. 2018). Finally, predictions will ultimately depend on the similarity of phenological cues between consumer and resource, and how they will change under climate change scenarios (Chmura et al. 2018).

**Towards robust forecasting of phenological mismatch**

While the Cushing hypothesis provides a testable hypothesis for predicting phenological mismatch in response to climate change, it is uncertain how much support there is for it. Here we have suggested two major reasons for this, both of which provide a clear path forward for progress in the field (Figure 2). First, most studies do not actually provide strong tests of the Cushing hypothesis because (a) studies do not collect equivalent fitness and population data on the consumer and resource, and (b) multiple mechanisms are not tested. This makes it difficult to refute the Cushing hypothesis if no evidence is found. Unless researchers are extremely clear about the mechanistic hypotheses they are testing, progress will continue to be slow. Second, pre-climate change baselines are not defined in the majority of studies in this literature. Lack of a relevant baseline makes it difficult to determine whether recent phenological mismatch can be attributed to climate change and limits our ability to understand and predict the direction and magnitude of phenological mismatch due to climate change.

Moving the field of phenological mismatch forward (i.e., from documentation towards forecasting) requires higher quality data that enables the testing of fundamental hypotheses and defining of key baselines. Below, we highlight how new approaches can improve testing of the Cushing hypothesis. These approaches will also provide the framework for building the required depth in evidence across studies to determine general quantitative patterns in phenological mismatch and their underlying mechanisms, an approach not currently possible given the state of the field.

*(i) Testing fundamental theory*

By having the potential to test multiple mechanisms and define key baselines, the integration of long-term data (i.e. that start before the1980s) and an experimental approach in a single system represents the ‘gold standard’ to move the field of phenological mismatch forward. However, in the absence of long-term data, the integration of observational data across a spatial gradient in climatic conditions, and thus variation in the relative timing of the interaction, with an experimental study can provide a powerful approach to test multiple hypotheses. For example, Samplonius et al. (2016) combine descriptive and experimental data to test many confounding factors and key assumptions of the Cushing hypothesis in their system (e.g., ontogenetic dietary shifts, habitat quality). Consequently, they advanced the field by providing a strong test of the Cushing hypothesis, demonstrating that this hypothesis is pertinent in their system, and ruling out competing hypotheses related to habitat quality.

Independent of whether observational data is available, experiments can be used to test and disentangle multiple mechanisms, for example direct from indirect effects or abiotic from phenological mismatches(e.g., Bauerfeind and Fischer 2013; Rudolf and Singh 2013). Experiments provide a high degree of inference about mechanism that descriptive data alone cannot match. For example, Berger et al. (2014) manipulated water temperature, stratification depth and the presence/absence of zooplankton in lake mesocosms to separate the direct effects of light and temperature on spring plankton dynamics from effects mediated through the consumer. They showed that temperature had little effect on zooplankton phenology and performance suggesting that climate change is unlikely to result in phenological mismatch in this system. The extrapolation of their results to the lake ecosystem are uncertain, however, given the lack of long-term data in the system and the uncertainty about phenological cues for *Daphnia* in the field (Berger et al. 2014).

Baselines themselves can also be used to test multiple mechanisms. Researchers with long-term data can define specific hypotheses about the expected vs. observed patterns of environmental change in their systems and then attribute variation in phenological mismatch to the main driver. For example, many lakes experienced large changes in nutrient inputs in the 1960s and 1970s (e.g., George 2012; Lake Washington (Edmondson 1994)), before the onset of recent climate change in the early 1980s. Researchers can test different breakpoints with their long-term data to determine where the non-linearity most likely began, thus providing evidence for which factor is likely to be the major driver of changes in phenological synchrony. For example, studies have shown that changes in nutrient inputs have had larger impacts on plankton dynamics in Lake Washington than warming over the past three decades (Law et al. 2009; Francis et al. 2014).

Higher temporal resolution data in some systems may help with teasing apart different mechanisms and testing the assumptions of the Cushing hypothesis. Regardless of aim, ideally, sampling frequency should be relative to the life history of the species of interest. For example, testing any mechanism related to life history theory requires data about ontogeny. In the great tit-winter moth system, caterpillar biomass sampling is usually conducted a few times a week (e.g., HMK041). Given that the larval stage of the winter moth life cycle is typically 4-6 weeks (Holliday 1977; Tikkanen et al. 2000), this sampling frequency allows researchers to obtain an accurate estimate of growth and development for this life cycle stage. In the lower trophic levels of aquatic systems, where there is often quick turnaround between producers and consumers, it can be difficult to determine how strong of an influence producers have on consumers and vice versa (Kitchell and Carpenter). Without higher temporal resolution data, it is difficult to assess whether this key assumption of the Cushing hypothesis (i.e., the dependence of the consumer on the producer) is supported. In cases where there are multiple short, potentially overlapping, generations, determining the seasonal order of the consumer vs. producer is necessary to assess the degree of dependence of the consumer on the producer.

Higher temporal resolution data will also help determine the seasonal availability of the consumer and resource, the other key assumption of the Cushing hypothesis (Figure 1b-d). For example, temperate aquatic and terrestrial systems differ in how many organisms survive the winter. While many terrestrial plants and herbivores have dormancy phases during the winter, in aquatic systems some zooplankton are never in complete dormancy (i.e., they remain at low densities), and thus do not meet the assumption of the hypothesis. Without more frequent data, accurately estimating the seasonal distribution of the lower trophic levels of some aquatic systems will be difficult. Nevertheless, some aquatic studies have been successful in achieving daily sampling frequency (HMK016, HMK036, HMK039), suggesting that—though more challenging—critical assumptions of the Cushing hypothesis can be tested in aquatic systems.

*ii) Defining baselines*

Researchers with long-term data have the potential to test whether synchrony vs. asynchrony was the pre-climate change baseline (Figure 2; Figure 3). By testing for a clear peak in fitness across years in relation to the relative timing of the interaction (i.e. evidence for the synchrony hypothesis) in the pre-climate change data, researchers can place themselves on the Cushing curve (Figure 3). Whenever possible, incorporating the most energetically expensive part of the consumer’s phenology or estimating when the measured phenological phase occurs relative to the most expensive part (e.g. food demands of great tit (*Parus major*) nestlings are highest 21 days after egg laying (i.e. the phenological phase most commonly monitored); Reed et al. 2013) will help define a peak in fitness. If there is no clear peak in fitness, then well-designed experiments (in systems where possible) that have the ability to reveal the limits of the curve (i.e. where consumer fitness reaches zero) can be used to reject the hypothesis of an asynchrony baseline. Knowledge of when/if different resources are available (e.g. Samplonius et al. 2016) and whether the interaction type changes throughout the season (add example from meta-analysis) can also help describe the limits of the curve.

Where long-term data are not available, there is no perfect solution to the challenge of defining a baseline. Null modeling to simulate historical data may be an alternative (e.g., Dornelas et al. 2014; Sgardeli et al. 2016; Kharouba et al. 2018). This null model could be used to put the magnitude of observed changes in phenology and performance since significant climate change began into context and to quantify how much variation may be due to noise (i.e., factors other than climate change; Kharouba et al. 2018). When not all required data is available across a long time period, the integration of experimental and observational data could here again be a useful approach (Figure 4). For example, in one study the performance data for an insect herbivore (winter moth; *Operophtera brumata*), and its main host plant (common oak; *Quercus robur*) are not available beyond X number of years from X location; however, researchers manipulated the timing of larval emergence relative to bud break to test for support for the Cushing hypothesis (Figure 4). Although the strength of the relationship may vary across populations (e.g. great tit work), a synthesis of the findings from the two approaches can shed light on the state of our knowledge of the system and can guide future research. Here we learned where the interaction is currently placed along the curve, how little of the curve is sampled with a 9 year post-climate change time-series, and that—without a clear baseline—it is difficult to predict how the performance of *O. brumata* will be affected by changes in phenological synchrony. Given the knowledge about the Cushing curve for this interaction, modelling pre-climate change baselines based on knowledge of phenological cues using a hindcasting approach (Senner et al. 2016; Deacy et al. 2017), process-based phenological models (Chuine and Régnière 2017), and/or phenologically explicit consumer-resource models (Bewick et al. 2016) could help to fill in data about the interaction before climate change began. Experiments could help define the parameters of a process-based model which could then elucidate which mechanisms may—or may not—appear feasible for the interaction (Chuine and R annual review).

*Final thoughts on forecasting*

Accurately forecasting phenological mismatch in response to climate change is a major test of ecological theory and methods. Previous work has identified factors that will help predict *when* a mismatch is more likely to lead to a population decline but this work is largely theoretical. While we have outlined here how to work around data limitations, the best tests of mismatch theory will come from data-rich systems where the energetic links between consumer and resource species are well established and understood. With such information in hand, researchers can then begin to consider forecasting longer-term demographic responses due to mismatch with continued climate change. Forecasting of phenological synchrony can be built off climate projection forecasts whenever phenological cues for both species (Chuine and Régnière 2017; Chmura et al. 2018) are well known. Forecasting longer-term demographic responses will require data about lifetime fitness for the consumer and comprehensive data on the resource.

As more information is gathered across habitats and diverse consumer-resource interactions, forecasting should eventually move beyond a system-specific approach and towards forecasting diverse systems through a multi-level modeling approach based on our mechanistic understanding of what drives mismatch. Such a model would be built on mechanistic studies of the mismatch hypothesis that carefully measured the fitness, phenologies, interactions and other still-unknown critical attributes of systems. The model would then allow researchers to measure currently unknown but critical attributes of species, sites, and ecosystems that control the shape, strength and uncertainty surrounding the Cushing curve. But progress towards such an approach requires the ability to combine knowledge from across diverse systems into one complex but well-defined curve. As we have outlined above, most current data in the field makes it hard or impossible to compare across systems. Given intrinsic differences between aquatic and terrestrial systems, progress may be accelerated if these systems are considered separately before being combined. Nevertheless, without improved methods – to test multiple mechanisms and define pre-climate change baselines – the goal ofgeneral predictions of the ecological consequences of shifts in phenological synchrony will remain well out of reach.

**Acknowledgements**

We thank Kathryn Cottingham, Mary O’Connor, Johan Ehrlen, Kjell Bolmgren and Steve Travers for interesting discussions. HMK thanks the professor writing retreats offered through the Centre for Academic Leadership at the University of Ottawa for support in writing this manuscript.

**References**

Gruner et al. 2008, A cross-system synthesis of consumer and nutrient resource control on producer biomass;

Borer et al. 2005, What determines the strength of a trophic cascade?

Edmondson 1994. Sixty years of Lake Washington: A curriculum vitae. Lake reserve. Manage, 10:75-84.

Hampton1, S.E., Scheuerell, M.D. and Schindler, D.E., 2006. Coalescence in the Lake Washington story: Interaction strengths in a planktonic food web. *Limnology and Oceanography*, *51*(5), pp.2042-2051.

Thackeray, S.J., 2012. Mismatch revisited: what is trophic mismatching from the perspective of the plankton?. *Journal of plankton research*, *34*(12), pp.1001-1010.

**Box 1. Glossary**

**Phenological synchrony-** changes in the relative timing of key life history events among pair-wise interacting species. Importantly, this term does not incorporate fitness consequences for either species.

**Phenological mismatch**- decreases in consumer fitness associated with changes in the relative timing of key life history activities among interacting species. This mismatch occurs between interacting species; it does not occur with the abiotic environment or due intraspecific processes.

**Cushing match-mismatch hypothesis**- this hypothesis postulates that there should be selective pressure for the consumer to temporally ‘match’ the peak of its energetic phase (i.e. reaching maximum fitness) with the peak of resource availability and if there is any change to the relative timing of the interaction, there will be a decrease in fitness for the consumer (i.e., a mismatch). At the curve’s limits, consumer fitness should fall to zero when the change in relative timing is sufficiently large. (Figure 1)

**Baseline-** a representative or benchmark time series of conditions that describes the species interaction. A baseline incorporates the historical range of variability in the relative timing of the interaction.

**Synchrony baseline-** a hypothesis that the pre-climate change baseline is that the most energetically demanding phase of the consumer is at the same time of peak resource availability, and thus fitness was at its maximum (i.e. a match).

**Asynchrony** **baseline-** a hypothesis put forward by Singer and Parmesan (2010) that postulates that, in some contexts, the pre-climate change baseline is one where there is no match (i.e., the most energetically demanding phase of the need not be at the time of peak resource availability).

**Stationarity-** any stochastic process with a constant underlying probability distribution

**Life-history theory-** the study of the causes and consequences of the diversity of life history strategies based on evolutionary theory.

**Food web theory-** the study of the patterns, causes and consequences of food-web structure.

**Tables**

Table 1. A comparison across studies of the type of performance data collected for consumer and resource. We define a life-history study as one that collected data at the individual level and a food-web study as one that collected data at the population or community (i.e., across species) level*.* Counts in the table are numbers of individual pair-wise interactions (n=46).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | Resource performance | | | |  |
|  | Life-history | Food-web | |  |
| None | Individual | Population | Community | *Totals* |
| Consumer  performance | Life-history | Individual | 6 | 0 | 4 | 18 | *28* |
| Food-web | Population | 1 | 1 | 7 | 7 | *16* |
| Community | 0 | 0 | 1 | 1 | *2* |
|  | | **Total** | **7** | **1** | **12** | **26** | **46** |

Table 2. A comparison across studies of the type of performance data collected for the consumer across systems and taxonomic group. Counts in the table are numbers of individual pair-wise interactions (n=46).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | System | | Taxonomic group | | | |
| Invertebrate | Vertebrate | | |
| Aquatic | Terrestrial | Fish | Bird | Mammal |
| Level of consumer  performance | Individual | 5 | 23 | 1 | 2 | 19 | 6 |
| Population | 14 | 2 | 9 | 4 | 2 | 1 |
| Community | 1 | 1 | 2 | 0 | 0 | 0 |
| **Sub-totals** | **20** | **26** | **12** | **6** | **21** | **7** |
|  | **Totals** |  | **46** |  |  |  | **46** |

**Figures**

Figure 1. Conceptualization of the Cushing match-mismatch hypothesis represented by the curve in panel (a). The hypothesis postulates that a consumer should temporally ‘match’ the peak of its energetic phase with the peak of resource availability (a,c) and if there is any change to the relative timing of the interaction (e.g., because the consumer shifts its phenology earlier (b) or later (d) relative to the resource), there will be a decrease in the consumer’s fitness (i.e., a mismatch) (a). At the curve’s limits, consumer fitness should fall to zero when the change in relative timing is sufficiently large (a). Curves in panels b-d represent the seasonal changes in the abundance of the consumer (red) and resource (black) where during some part of the year abundance declines to zero.

Figure 2. A simplified flow diagram for forecasting climate change effects on consumer fitness as predicted by the Cushing hypothesis. First, both major assumptions must be met; if, for example, consumer density is the major controller on its own fitness, then no further work is warranted. If both assumptions are met, forecasting requires both knowledge of the ultimate mechanisms (which often relate to testing assumption 1) and the pre-climate change baseline. In some cases, the mechanism predicts the pre-climate change baseline (i.e., in both our examples).; we color mechanisms in whether they predict both pre-climate change synchrony and asynchrony depending on specifics (purple), synchrony (blue) or asynchrony (red). Note that multiple mechanisms may operate in many systems, and that tin no way, serves

Figure 3. Conceptualization of key assumptions, and resulting implications for climate change predictions, often made about the Cushing hypothesis when pre-climate change baselines are not defined. (a) Differences in the phenological time-series of a consumer-resource interaction, where red represents the resource and black represents the consumer, during conditions of stationarity,; when the environment becomes non-stationary (shown here just after 1980) the consumer and resource each can shift in varying directions, leading potentially to shifts in synchrony. (b) Most studies in the current literature assume that consumer fitness was highest before climate change (i.e., a match; synchrony hypothesis). However, an alternative hypothesis put forward by Singer and Parmesan (2010) (i.e., the asynchrony hypothesis) postulates that conditions before climate change may not represent a ‘match’ in the system. The implications for climate change predictions for the two hypotheses are illustrated: If the synchrony baseline hypothesis is supported, then climate change will necessarily lead to declines in consumer fitness. If the asynchrony baseline hypothesis is supported, climate change may not lead to declines in consumer fitness. (c) Without establishing a pre-climate change baseline and defining where an interaction falls along a curve, it is difficult to predict how climate change-driven changes to the relative timing of the interaction may affect consumer fitness. For example, in scenario 1, climate change could lead to an increase or decrease in consumer fitness depending on how the relative timing of the interaction changes. Similarly, in scenario 2 (i.e. alternative asynchrony baseline), climate change could lead to an increase but to a much smaller degree.

Figure 4. Case study demonstrating the integration of experimental (a) and observational data (b) relating to the Cushing hypothesis in a single system- the winter moth (*Operophtera brumata*) and oak (*Quercus robur*). (a) Experimental raw data was obtained from Tikkanen and Julkunen-Tiitto (2003) and result from two experiments: in the first experiment, the authors manipulated the number of days that neonates (i.e. early instar larvae) spent without food (green points). In the second experiment, they manipulated the emergence times of larvae (red points). There were four cohorts, each separated by intervals of 3-5 days. All *O. brumata* eggs originated from laboratory stock originally from Turku, Finland whereas the foliage originated from trees near Banchory, NW Scotland. (b) Inter-annual variation in relative timing between median egg hatch date of *O. brumata* and the median bud opening date of *Q. robur* from 1996-2005 in the Netherlands. Horizontal error bars represent the lower and upper quartiles of the data. Raw data from the observational study was retrieved from VanAsch and Visser 2007 Figure 2. In this system, negative values along the x-axis denote where egg hatching occurred before bud opening, whereas positive values indicate egg hatching occurred after bud opening.

Figure 1.



Figure 2.

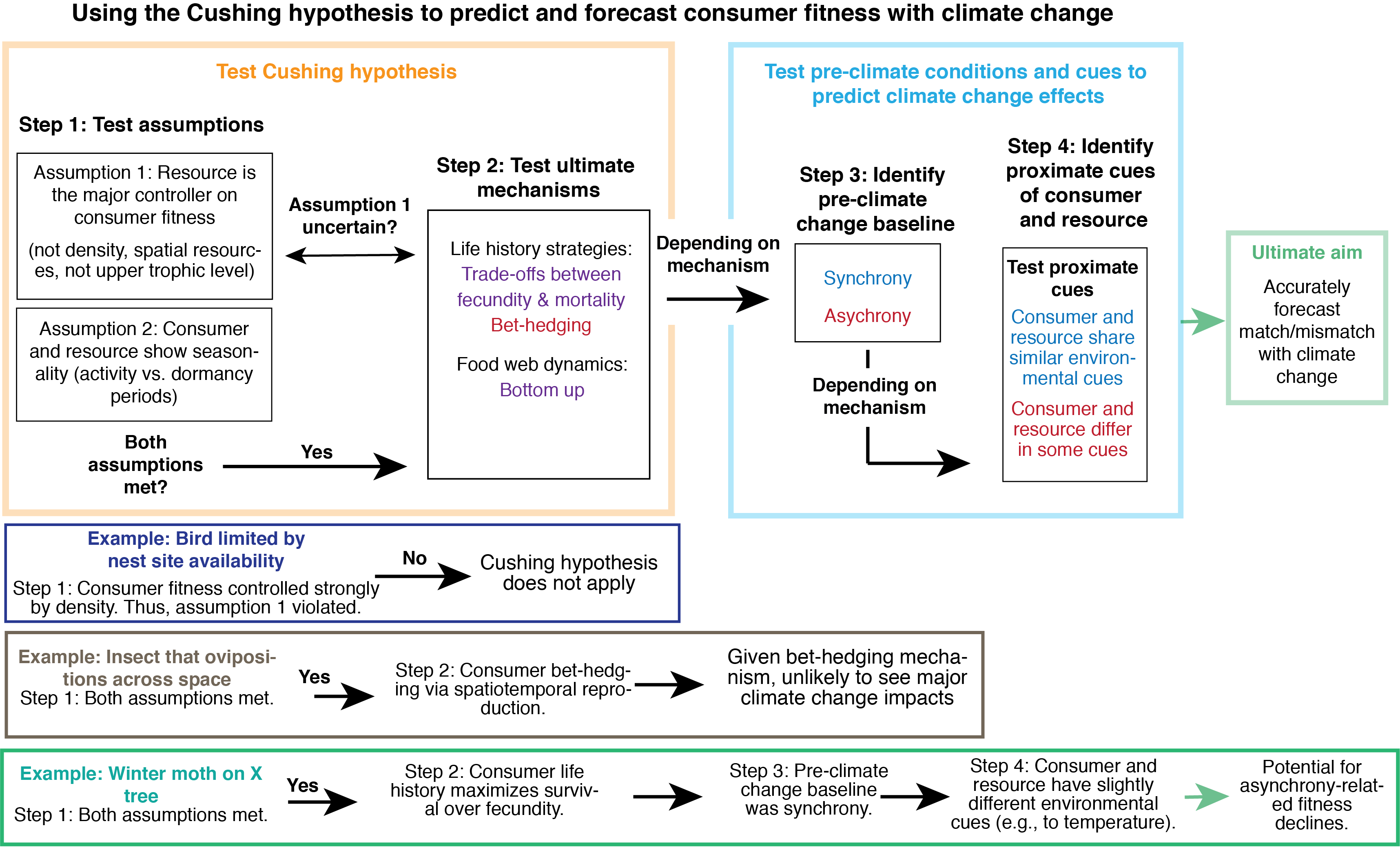


Figure 3.



Figure 4.



**Appendix**

*Literature search*

We located papers relating phenological data from trophic interactions to fitness and/or performance of the consumer and/or the resource by conducting keyword searches in ISI Web of Science published up to June 2017. Keywords included phenolog\* AND mismatch\* OR synchron\* AND interact\* AND (fitness\* OR performance\*). Our initial search netted 2906 papers so we further refined our search by excluding categories that included engineering and computer science. This resulted in 393 papers. From these, we focused on observational studies and excluded studies if they did not: (1) measure phenology directly (e.g. used derived measures of phenology, such as NDVI or spring temperature); (2) measure phenology of at least one of the species; (3) quantitatively link consumer fitness or performance to the relative timing between consumer and resource; and (4) explicitly state that the two species interacted (e.g. specifying type of interaction). Estimates of consumer fitness or performance (i.e., growth or development) had to be direct estimates (e.g. we did not include studies that used diet proportions to measure consumer performance) and included a range of proxies, for example, growth, survival, abundance. To ensure a reasonable sample size and to include studies across different major biomes, we included interactions that were resolved to the family-level and below. To reduce redundancy within systems, we only kept studies that were unique across pairs-location-year combinations. If more than one measure of phenology was included, we chose the one used by the authors to calculate mismatch and examine its impact on performance.

Our final review included 43 studies with 46 pair-wise species interactions (3 studies had 2 interactions). These studies encompassed terrestrial, marine and freshwater ecosystems as well as a large latitudinal gradient (map?).

*Summary of studies and interactions*

We classified studies as ‘climate change’ or `fundamental’ depending on whether they mentioned climate change in the abstract or in the introduction of the paper. Based on the type of data collected for the consumer and resource, we classified these studies as life history (i.e. one that collected data at the individual level) or food web-based (i.e. one that collected data at the population or community (i.e., across species); Box 1). To determine whether studies had the potential to define pre-climate change baselines, we measured the study’s time span and years of data based on the years where phenology data was available for both the consumer and resource, and consumer performance data was available.