RESEARCH ARTICLE

Phylogenetic analysis of adaptation in comparative physiology and biomechanics: overview and a case study of thermal physiology in treefrogs

Daniel S. Moen[1,‡], Elisa Cabrera-Guzmán[1], Itzue W. Caviedes-Solis[2,3,*], Edna González-Bernal[4] and Allison R. Hanna[1]

ABSTRACT

Comparative phylogenetic studies of adaptation are uncommon in biomechanics and physiology. Such studies require data collection from many species, a challenge when this is experimentally intensive. Moreover, researchers struggle to employ the most biologically appropriate phylogenetic tools for identifying adaptive evolution. Here, we detail an established but greatly underutilized phylogenetic comparative framework – the Ornstein–Uhlenbeck process – that explicitly models long-term adaptation. We discuss challenges in implementing and interpreting the model, and we outline potential solutions. We demonstrate use of the model through studying the evolution of thermal physiology in treefrogs. Frogs of the family Hylidae have twice colonized the temperate zone from the tropics, and such colonization likely involved a fundamental change in physiology due to colder and more seasonal temperatures. However, which traits changed to allow colonization is unclear. We measured cold tolerance and characterized thermal performance curves in jumping for 12 species of treefrogs distributed from the Neotropics to temperate North America. We then conducted phylogenetic comparative analyses to examine how tolerances and performance curves evolved and to test whether that evolution was adaptive. We found that tolerance to low temperatures increased with the transition to the temperate zone. In contrast, jumping well at colder temperatures was unrelated to biogeography and thus did not adapt during dispersal. Overall, our study shows how comparative phylogenetic methods can be leveraged in biomechanics and physiology to test the evolutionary drivers of variation among species.

INTRODUCTION

Experimental studies in comparative physiology and biomechanics often demonstrate the functional significance of phenotypes in a given environment. A close fit between environment and function can be seen as indicating that a phenotypic state is adaptive, with the assumption that the trait and its function increase the organism's fitness in its current environment (Reeve and Sherman, 1993). Such studies, sometimes called the equilibrium approach, can be and frequently are performed on single species (Lauder, 1982, 1996).

Demonstration of adaptation across species is much less common in comparative physiology and biomechanics (comparative physiology hereafter for brevity; Vogel, 2007). In explicitly historical studies, researchers collect data for multiple species and examine character evolution along a phylogeny. Studies analyzing morphological data with complex phylogenetic comparative analyses are widespread (Cooper et al., 2016a). However, similar studies using physiological data are much rarer owing to three key, interrelated challenges (Lauder, 1990, 2003; Garland et al., 2005; Bauer et al., 2020): (1) collecting data from multiple species distributed across broad geographic areas is expensive and logistically difficult; (2) challenging conditions limit the type of data that can be collected in the field; and (3) difficult experimental procedures using finely calibrated equipment may be hard to replicate on multiple species. Despite these challenges, such studies are essential for understanding the evolution of physiological diversity (Lauder, 2003), particularly over deep time scales.

Phylogenetic studies of adaptation have taken two key tracks. First, early practitioners emphasized studying the order of character change along a phylogeny (Lauder, 1982, 1991; Baum and Larson, 1991). Changes in phenotype that occurred along with or subsequent to changes in the environment on a branch indicated adaptation. This practice followed defining an adaptation as a character state that originated in response to natural selection in a novel environment (Lewontin, 1978; Gould and Vrba, 1982; Amundson, 1996). Yet this approach has largely fallen out of practice for many reasons. Ancestral states are hard to accurately estimate, particularly on phylogenies with only extant species (Cunningham et al., 1998; Cunningham, 1999; Losos, 1999). Moreover, in fast-evolving traits or on long branches, multiple changes on a single branch cannot be detected. Alternatively, infrequent evolutionary change in both phenotype and environment may limit robustly testing the fit between the two (Martins, 2000). Lastly, change in both phenotype and environment along the same branch may make it difficult to test which change was the cause of adaptation: did phenotypic change permit colonization of a novel environment, or did environmental change cause phenotypic adaptation (Lauder, 1991)? In the extreme case of change in several traits and environments along a single branch, causal inference may be impossible (Lauder, 1982, 1990; Leroi et al., 1994; Maddison and FitzJohn, 2015; Uyeda et al., 2018).

[1]Department of Integrative Biology, Oklahoma State University, Stillwater, OK 74078, USA. [2]Science Unit, Lingnan University, Hong Kong S.A.R., China. [3]Department of Biology, University of Washington, Seattle, WA 98105, USA. [4]CONACYT - CIIDIR Oaxaca, Instituto Politécnico Nacional, Santa Cruz Xoxocotlán, C.P. 71230, Oaxaca, México.
*Present address: Biology Department, Swarthmore College, Swarthmore, PA 19081, USA.

‡Author for correspondence (daniel.moen@okstate.edu)

D.S.M., 0000-0003-1120-0043; E.C.-G., 0000-0001-5590-649X; I.W.C.-S., 0000-0002-8440-3645; E.G.-B., 0000-0001-7817-4156; A.R.H., 0000-0001-7946-9781

List of symbols and abbreviations

BM	Brownian motion
CT_{min}	critical thermal minimum
L80	lower temperature threshold for 80% of peak jumping performance
MCC	maximum clade credibility
OU	Ornstein–Uhlenbeck
PGLS	phylogenetic generalized least squares
TPC	thermal performance curve
α	rate of approach to the primary optimum
θ	primary optimum in OU model
σ	scale of change in models of evolution
σ^2	rate of stochastic character change in models of evolution

A second, alternative phylogenetic approach tests the long-term association between phenotypes and environments, the classical comparative method (Harvey and Pagel, 1991; Martins, 2000). Approaches based on this association – including phylogenetic regression and ANOVA – are now increasingly common in comparative physiology and implicitly test for adaptation (Martins, 2000; Olson, 2021). However, the most common approaches primarily aim to avoid statistical problems in hypothesis testing, such as the non-independence of species (Hansen and Orzack, 2005; Hansen, 2014; Pennell, 2015). Such approaches do not directly model the process of adaptation of a phenotype to a particular environment (Hansen, 1997; Butler and King, 2004), which leads to statistical and interpretational problems. Hansen (1997) introduced a method, based on the Ornstein–Uhlenbeck (OU) process, that models evolution as stochastic movement towards an adaptively optimal state. This 'Hansen model' (Butler and King, 2004) has its foundations in earlier quantitative-genetic models of stabilizing selection, with long-term fit of a character state to its environment as positive evidence for adaptation (Hansen, 1997; Martins, 2000).

Despite its age, the Hansen model remains greatly underutilized. As work in comparative physiology becomes increasingly phylogenetic, we believe that understanding this approach and its advantages for studying physiological adaptation is timely. Thus, in this study, we first describe Hansen's (1997) OU model of adaptation in more detail. We next demonstrate its use in a case study of physiological evolution in treefrogs. We then discuss commonly misunderstood statistical properties of the model, challenges in implementation, and the effects of studying relatively few species. Finally, we provide data analysis tutorials to help researchers use the Hansen model in their own work.

Overview of the Hansen OU model of adaptation

The Hansen model has been described and discussed considerably in the evolutionary literature (see reviews in Hansen, 2014; O'Meara and Beaulieu, 2014). Yet proper interpretation of the model, as well as best practices for implementation, remain poorly understood (Hansen, 2014; Cooper et al., 2016a,b). We thus summarize the model here and distinguish it from more frequently used approaches. Throughout this discussion we use 'environment' to indicate a discrete selective factor (e.g. diet, habitat, biogeographic region); such factors are also commonly called selective regimes. We use 'phenotype' to indicate the value of a continuous trait that responds to selection in the environment.

Most phylogenetic comparative analyses model trait evolution under Brownian motion (BM; Garamszegi, 2014; Pennell, 2015; Cooper et al., 2016a). BM is a pure diffusion process, with trait change over time along phylogenetic branches, $dY(t)$, described by

the stochastic differential equation:

$$dY(t) = \sigma dB(t), \qquad (1)$$

where $dB(t)$ indicates a standard normally distributed change [i.e. $dB(t){\sim}N(0,1)$] and σ indicates the scale of change (Butler and King, 2004). Viewed differently, the squared scale of change, σ^2, is the rate of evolution (Martins, 1994; O'Meara et al., 2006): if changes are larger at each time step, more variation will accrue in a clade over time. Because this simplest version of BM has no tendency to move toward any given trait value, species are always expected to have values similar to those of their ancestors. Moreover, species are expected to be similar to each other in direct proportion to the amount of evolutionary history they share. Although BM can be consistent with adaptive evolution (Felsenstein, 1988; O'Meara et al., 2006; Harmon et al., 2010), it has no tendency to increase adaptive fit over time (Martins, 1994; Hansen and Martins, 1996).

The OU process shares some elements with BM but importantly adds deterministic change toward a central value. It has often been described as a rubber band model (Felsenstein, 1988): taxa that wander (stretch) too far from a central value are pulled back toward it, and the farther they evolve away from it (the more they stretch the rubber band), the stronger the pull to return to the center (the more strain energy stored in the rubber band). Here, changes in a phenotype $Y(t)$ between times t and $t+dt$ are modeled as:

$$dY(t) = -\alpha(Y(t) - \theta)dt + \sigma dB(t), \qquad (2)$$

where dt is a infinitesimally small unit of time, θ is a central 'optimal' value and α is the strength of the pull toward the optimum (Butler and King, 2004). The right-hand side of the equation is the same as in the BM model, meaning that when $\alpha=0$ (i.e. no adaptation toward the optimum), the OU process collapses into pure BM (Eqn 1).

When introduced for comparative biology, the OU process was originally described as a model of stabilizing selection in which α is the strength of selection and the BM diffusion represents genetic drift (Martins, 1994; Hansen and Martins, 1996). Although this population-genetic interpretation has intuitive appeal, macroevolution proceeds far too slowly for the OU parameters to be interpreted this way (Lynch, 1990; Hansen, 1997, 2012; Harmon et al., 2010; Uyeda et al., 2011). Thus, Hansen (1997) described θ as the primary optimum for an environment: the phenotype species in that environment would have if no other factor affected the focal trait's evolution. In contrast, species can be considered to have (and already be at) a species-specific optimum value for the focal trait. These latter optima differ from species to species because the focal trait is likely also under selection for other species-specific functions (Hansen, 1997), as implied by widespread trade-offs (Roff and Fairbairn, 2007). Moreover, other factors may constrain evolution toward the primary optimum, including correlation with other traits or genetic and developmental constraints (Hansen, 1997). Thus, even though all species are pulled toward the primary optimum for their current environment, their phenotypes may not take that optimal value. Importantly, many sources of deviation from the primary optimum are likely shared by closely related species. Thus, even under the OU process, closely related species are expected to deviate from the primary optimum similarly. This is the role played by the BM diffusion on the right-hand side of Eqn 2. Overall, another way to think about the OU process is that each lineage in the same environment is a somewhat independent evolutionary replicate adapting toward the same primary optimum (which we refer to as simply 'optimum' hereafter for brevity). All of these lineages deviate in their own

idiosyncratic way, associated with other aspects of their biology that are shared with other species to varying degrees (Hansen, 1997).

The heart of the OU process as a model of adaptation is the idea that the current phenotype of a species results from an evolutionary history of adaptation to different environments. A consequence is that past adaptation may have residual effects on current phenotypes. Adaptation is not expected to happen instantaneously (Labra et al., 2009; Hansen, 2012), so when a species changes environment, it may take time to adapt toward the new primary optimum (Moen et al., 2016; Toljagić et al., 2018). The parameter α represents the rate of movement toward the optimum. Moreover, a more intuitive understanding of α can result from reparameterizing it as $\log(2)/\alpha = t_{1/2}$, the phylogenetic half-life (Hansen, 1997; Hansen et al., 2008). This half-life indicates the expected time a species would take to adapt halfway from its current phenotype toward the optimum. If α is large, the pull of the optimum is strong, and species adapt quickly. The half-life has units of time, so whether a half-life is considered small or large partly depends on the length of the phylogeny studied. For example, a half-life of 10 million years (myr) means slow adaptation on a phylogeny 1 myr long but rapid adaptation on a 100-myr phylogeny.

As a concrete example, imagine studying evaporative water loss (EWL) rate in a lizard species that lives in a tropical dry forest. Phylogenetic analysis suggests that the species occurred in tropical rainforests earlier in its evolutionary history. Moreover, its current EWL rate is approximately 75% of the distance from a typical rainforest EWL rate (i.e. the rainforest optimum) towards a typical dry-forest EWL rate (i.e. the dry-forest optimum). This result could stem from an early shift to the dry forest coupled with a weak pull toward the dry-forest optimum (Fig. 1A). Alternatively, the species may have colonized the dry forest more recently with a strong pull toward the dry-forest optimum (Fig. 1B). Regardless, the history of adaptation of a species may result in a phenotype that is not close to the primary optimum of its current environment (Moen et al., 2016; Toljagić et al., 2018). Interestingly, in this way the OU process helps explain variation among species adapting to the same environment (Fig. 1). It also accommodates the expectation that evolutionary history matters when interpreting adaptation (Lauder and Liem, 1989; Lauder, 1990).

This explicit modeling of adaptation to differing environments throughout the history of a species is the major difference between the Hansen model and BM. This difference leads to two key interpretational problems when using BM in studies of adaptation. First, even when species change environments throughout their history, they will never be modeled as better adapted to one versus the other. BM only allows correlated changes between a phenotype and its environment, rather than phenotypic tracking to changes in environment (Hansen and Orzack, 2005; Hansen et al., 2008). So if the phenotype of a species is far from the optimum of its current environment and then the species changes environment, its phenotype will be modeled as equally far from the optimum of its new environment. This would occur even if a species were already close to the optimum of its new environment, which paradoxically may have been the cause of changing environment in the first place (i.e. high fitness in an environment may promote colonization of that environment). This property of BM has been called inherited maladaptation (Hansen and Orzack, 2005; Hansen et al., 2008).

Second, common approaches that often use BM to model species non-independence, such as independent contrasts (Felsenstein, 1985) and phylogenetic generalized least squares (PGLS; Martins and Hansen, 1997), only incorporate the phylogeny in terms of residual variation around a fitted relationship, not in estimating the relationship itself (Martins and Hansen, 1997; Hansen and Orzack, 2005). Many researchers do not distinguish this important detail, instead expecting that any kind of similarity among species should be accounted for in a phylogenetic analysis (Revell, 2010; Hansen and Bartoszek, 2012). Yet the phenotypes of closely related species may also be similar because they occur in and are adapting to a similar (or the same) environment as a shared common ancestor (Labra et al., 2009). Such inherited similarity in environment should
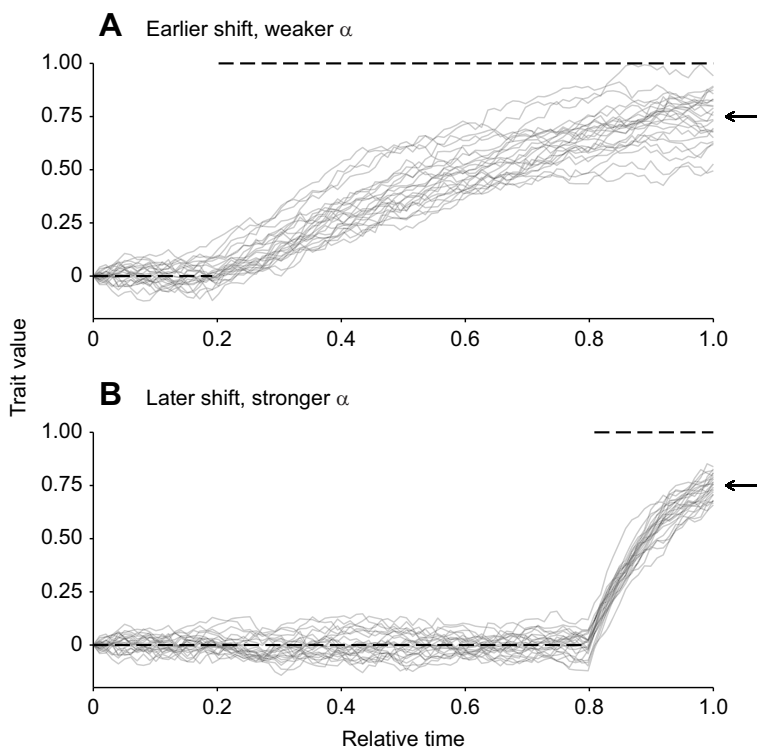

Fig. 1. Simulation example of three ways the Ornstein–Uhlenbeck (OU) process can explain the distinction between a primary optimum (θ) and species' values (their own optima). In A, species start in an environment whose primary optimum is 0 (lower dashed line). After 20% of their history, they shift to a new environment with a primary optimum of 1.0 (upper dashed line). Based on the α chosen for simulation, species' phenotypes are expected to be near 0.75 (the arrow) after the next 80% of their history. Viewed alternatively, the phylogenetic half-life, $\ln(2)/\alpha$, is 0.4: species move on average half the distance from 0 to 1 (i.e. 0.5) after 0.4 units of time, then another half (from 0.5 to 0.75) in the next 0.4 units. Variation across lines represents stochasticity in the process, as represented by 25 simulation replicates, each with their own distinct history. In B, species adapt to 0 for the first 80% of their history, then shift to the new environment and quickly move toward 1.0 (i.e. α is larger, with a half-life of 0.1). Both A and B have the same stochastic parameter value (σ^2), but variance is higher in A because both σ^2 (same in A and B) and α (higher in B) affect the variance of the OU process (Hansen, 1997). These two graphs thus show that individual species optima may differ from the primary optimum of their current environment owing to (1) slow adaptation despite a long history in the environment (A), (2) a short history in an environment despite fast adaptation (B) or (3) unmodeled factors, represented by the stochastic component (both A and B). In practice, all three factors are likely to be at play (Hansen, 1997, 2014; Moen et al., 2016; Toljagić et al., 2018). Note that when one considers multiple species, the stochastic component representing unmodeled factors will be somewhat shared among closely related species.

not be removed by a statistical model (Hansen and Bartoszek, 2012; Taylor and Thomas, 2014). Doing so causes a misfit of the adaptive relationship (Hansen and Orzack, 2005; Hansen et al., 2008; Labra et al., 2009) and high Type-I error rates (Revell, 2010). OU methods explicitly account for this shared adaptation, as does the common approach of using PGLS to simultaneously estimate a regression line and phylogenetic signal (Revell, 2010; Hansen and Orzack, 2005). However, the latter approach still suffers from the problem of inherited maladaptation.

Overall, the Hansen model solves many of the problems of improper modeling of the biological process of adaptation. BM-based models are simpler to implement, but they tend to focus on statistical fixes and do not properly account for the adaptive process. We have described the Hansen model in its simplest form for testing adaptation: environments differ in primary optima (θ) but not in rate of approach to the optima (α) or rate of stochastic evolution (σ^2). Likewise, we demonstrate the model's use for studying adaptation in this framework. However, more recent extensions to this model include continuous environments (Hansen et al., 2008), ANCOVA-like designs (Escudero et al., 2012), distinct α and σ^2 across environments (Beaulieu et al., 2012), and multivariate applications (Bartoszek et al., 2012). Moreover, these methods are provided in a wide array of R packages that vary in their implementation details and options (O'Meara and Beaulieu, 2014). We return to these extensions, as well as limitations and challenges of the Hansen model, in the Discussion.

Case study: thermal physiology in hylid treefrogs

The latitudinal biodiversity gradient is a pattern of declining species diversity from tropical to temperate latitudes (Willig et al., 2003; Hillebrand, 2004; Mittelbach et al., 2007; Mannion et al., 2014; Pontarp et al., 2019). A key hypothesis for explaining this gradient is that species have rarely dispersed from the tropics to the temperate zone because of ecological niche conservatism (Farrell et al., 1992; Ricklefs and Schluter, 1994; Wiens and Donoghue, 2004), meaning lineages do not adapt to novel environments outside their native range. Species distribution models of geographic range limits have shown that temperature seasonality and cold extremes are the key climatic factors limiting colonization of the temperate zone (Willig et al., 2003; Wiens et al., 2006). However, the physiological constraints that cold temperatures impose are often unclear (Chown et al., 2004; Gaston et al., 2009; Bozinovic et al., 2011; Spicer et al., 2019).

Thermal tolerance (e.g. surviving cold temperatures) and the effects of temperature on functional performance (e.g. in locomotion) are two strong candidates for explaining how physiology may limit colonization of the temperate zone (van Berkum, 1988; Bozinovic et al., 2011). First, tolerance to low temperatures predicts northern range limits in thermally sensitive species (Calosi et al., 2010), and terrestrial ectotherms at higher latitudes and cooler climates tolerate colder temperatures (Snyder and Weathers, 1975; Kimura, 2004; Sunday et al., 2011, 2019; Bennett et al., 2021). Thermal tolerance can be represented by critical thermal minima and maxima, which are the lowest and highest temperatures, respectively, at which an organism loses locomotor ability (Lutterschmidt and Hutchison, 1997). Thermal minima and maxima are often close to absolute (i.e. lethal) limits and have been called the point of 'ecological death': if an organism cannot move, then its likelihood of survival drastically decreases (Cowles and Bogert, 1944; Huey and Kingsolver, 1989; Sunday et al., 2011). Lower thermal tolerances have been shown to be more evolutionarily labile than upper tolerances (Araújo et al., 2013;

Bennett et al., 2021), and physiological traits in general tend to show low phylogenetic signal (e.g. Blomberg et al., 2003; Hertz et al., 2013; Krause et al., 2014). These patterns may both suggest adaptation in lower tolerances. For example, low phylogenetic signal in tolerances (e.g. von May et al., 2017) may result from adaptation to a selective factor (e.g. elevation) that itself shows relatively low signal. This is a situation where the Hansen model excels (Labra et al., 2009; Kozak and Wiens, 2010).

Second, in ectotherms, locomotor performance must be maintained at high levels despite varying body temperatures (John-Alder et al., 1988; van Berkum, 1988; John-Alder et al., 1989; Bozinovic et al., 2011). For example, high-latitude species have lower minimum field body temperatures (John-Alder et al., 1988), and such species and populations often perform better at cold temperatures than species from lower latitudes (John-Alder et al., 1988; Wilson, 2001; Li et al., 2018). Additionally, the breadth of temperatures at which organisms perform well usually increases with latitude (Navas et al., 2008), consistent with the increase in temperature variation at higher latitudes (Sunday et al., 2011). These results suggest that selection favors broad performance curves in temperate ectotherms in order to function well at cold temperatures (John-Alder et al., 1988; van Berkum, 1988), which may consequently affect biogeographic dispersal.

Treefrogs of North and South America (Amphibia: Anura: Hylidae) are an excellent group for addressing physiological evolution associated with the latitudinal diversity gradient. They occur at both high and low latitudes (temperate and tropical regions), with most species in the tropics (Duellman, 1999). Moreover, five hylid clades have northern range limits in Mexico, yet only one of them – Hylini, also known as the Middle American clade – has colonized the temperate zone (Smith et al., 2005, 2007; Wiens et al., 2006; Moen et al., 2009). By examining how physiology changed in Middle American hylids upon colonizing the temperate zone, we can infer what may have historically limited other clades from colonizing. Such factors may also be applicable to many other organisms.

Previous studies of thermal physiology and biogeography in anurans have shown broader thermal performance curves (TPCs) in cooler climates (Renaud and Stevens, 1983; John-Alder et al., 1988; Wilson, 2001). Moreover, critical thermal minima (CT_{min}) decrease as latitude increases (Brattstrom, 1968; Layne and Romano, 1985; John-Alder et al., 1988). Recent studies have examined the evolution of thermal tolerances in anurans (von May et al., 2017, 2019). However, the joint evolution of TPCs and CT_{min}, and particularly their role in historically limiting the transition of tropical anuran lineages into the temperate zone, has not been studied. More broadly, only one study has compared the evolution of both these characteristics as they relate to latitudinal differences in climatic niche, in *Drosophila* (MacLean et al., 2019).

Here, we examined the evolution of physiology in temperate and tropical hylid frogs. We estimated the CT_{min} of six temperate and six tropical species in the Middle American clade. We then estimated TPCs in jumping for each species and calculated the lower temperature (L80) at which the performance of each species significantly declined. We used the OU model of adaptation to test whether broad TPCs, higher cold tolerance, both or neither evolved when hylid frogs colonized the temperate zone from the Neotropics.

MATERIALS AND METHODS
Sampling

We sampled species evenly across the Middle American clade (Hylini) of the frog family Hylidae (Fig. S1; Wiens et al., 2005;

Faivovich et al., 2018). This clade is largely endemic to Middle America (Mexico to Panama), but it includes species in temperate North America, Asia and Europe. In the temperate zone, we collected six species from Oklahoma, Texas and Arkansas, USA, which included three early-spring breeders (*Pseudacris fouquettei*, *Pseudacris crucifer* and *Acris blanchardi*) and three late-spring breeders (*Hyla cinerea*, *Hyla avivoca* and *Hyla arenicolor*). These two clades independently colonized the temperate zone (Smith et al., 2005) approximately 57 and 44 million years ago, respectively (Jetz and Pyron, 2018). In the tropics, we sampled six species from Oaxaca, Mexico, in the municipalities of Santiago Comaltepec (*Ptychohyla zophodes*, *Charadrahyla nephila*, *Exerodonta abdivita* and *Smilisca cyanosticta*) and Pluma Hidalgo (*Smilisca baudinii* and *Tlalocohyla smithii*). These species represented six of the eight major clades of the Middle American clade (Fig. S1; Smith et al., 2007; Faivovich et al., 2018). We collected adult male frogs at night during the breeding season (temperate taxa: March–June and August 2017 and 2018; tropical taxa: June 2018; see Supplementary Materials and Methods). We sampled 5–7 individuals for most species (Table 1). These sample sizes are consistent with previous comparative studies of amphibian TPCs (John-Alder et al., 1988; Gvoždík and Van Damme, 2006).

We performed all work with appropriate collecting and animal ethics permits. Collection permits were provided by the Oklahoma Department of Wildlife Conservation (permit 5552719), Arkansas Game and Fish Commission (permit 032820171), Texas Parks and Wildlife Department (SPR-0416-112) and SEMARNAT México (SGPA/DGVS/004473/18). All work was done under Oklahoma State University ACUP AS-17-3.

Thermal performance curves
We examined TPCs in jumping, a frog's primary mode of locomotion (Gans and Parsons, 1966; Jenkins and Shubin, 1998; Mendoza et al., 2020). Different jumping variables such as velocity and acceleration have been shown to be positively correlated (Moen, 2019). Thus, we used peak velocity during take-off as the performance variable for generating TPCs, and we expect results to be similar for acceleration.

Following previous studies (John-Alder et al., 1988; Wilson and Franklin, 2000; Wilson, 2001), we collected jumping performance data at six different temperatures: 8, 14, 20, 26, 32 and 35°C. Though only 3°C separated our two highest temperatures, we selected the highest temperature based on previous anuran work on

TPCs (reviewed in Navas et al., 2008). Moreover, we were not able to measure performance for *S. baudinii* at 8°C, nor for *S. cyanosticta* at 8 and 14°C, as they refused to jump (see also Navas 1996 for similar behavior seen in other lowland tropical species). Thus, we tested them by increasing their body temperature by 1°C each trial until we found the minimum temperature at which they would jump (*S. baudinii* >12.5°C; *S. cyanosticta* >16°C).

We initially housed all animals individually in the laboratory at 20°C, a typical active temperature for all study species (Duellman, 2001). We kept them at 20°C for 1 week to minimize any potential effects of acclimation on performance (Dunlap, 1980; Renaud and Stevens, 1983; John-Alder et al., 1988, 1989; Whitehead et al., 1989; Wilson and Franklin, 2000), though little evidence exists for acclimation-based performance differences in anurans (see Discussion). We fed frogs fruit flies or crickets (based on gape size) every other day until starting data collection, then fed them once in the middle of the week of performance trials.

To change body temperature, we placed the frogs in a water bath at the desired temperature until their body temperature reached the treatment temperature, as determined with an infrared thermometer (see Supplementary Materials and Methods for additional detail, including demonstration that our methods did not induce thermal shock in animals). We randomly assigned experimental temperatures to each frog and tested one or two temperature treatments per day, waiting at least 5 h between trials at different temperatures. Trials generally lasted 1 week. If an individual did not perform well at a given temperature (e.g. it fatigued before we could obtain a video), additional days were added to retest that temperature. Even though performance may gradually decline when individuals are measured over the course of a week (Zug, 1985), we found similar performance at the beginning and end of our experiments (Supplementary Materials and Methods).

We recorded jumping performance using a Fastec TS5 high-speed camera, filmed lateral to the jump at 250 frames s^{-1} and with an exposure time of 500–1000 µs. We stimulated each frog to jump by gently tapping its back leg or clapping behind it. We recorded 3–4 jumps during a given set of trials, based on the number necessary to obtain peak performance in previous studies (Nauwelaerts et al., 2007; Moen et al., 2013, 2021b; Mendoza et al., 2020). We re-established the test temperature by placing frogs in the water bath between all jumps.

To estimate peak velocity, we digitized videos with ImageJ v.1.52a (Schneider et al., 2012). These digitized coordinates yielded

Table 1. Species means used in comparative analyses

Species	Details	CT$_{min}$ (°C)	n	L80 (°C)	n
Temperate					
Acris blanchardi	Early-breeding	0.27±1.16	5	10.23	7
Pseudacris crucifer	Early-breeding	0.64±1.26	5	12.28	7
Pseudacris fouquettei	Early-breeding	0.56±1.02	5	10.88	6
Hyla arenicolor	Late-breeding	5.44±0.31	8	16.46	7
Hyla avivoca	Late-breeding	2.86±0.26	5	14.78	9
Hyla cinerea	Late-breeding	3.55±0.08	6	16.22	7
Tropical					
Charadrahyla nephila	Highland	6.42±0.06	6	17.59	5
Exerodonta abdivita	Highland	6.33±0.27	6	10.67	6
Ptychohyla zophodes	Highland	5.14±0.23	5	12.70	7
Smilisca baudinii	Lowland	8.92±0.90	3	27.45	6
Smilisca cyanosticta	Lowland	6.90±0.18	5	13.74	7
Tlalocohyla smithii	Lowland	–	–	16.04	1

We only calculated the standard errors (s.e.m.) for critical thermal minimum (CT$_{min}$), given that the lower temperature threshold for 80% of peak jumping performance (L80) was determined from a curve estimated across all individuals of each species. Samples sizes (*n*) for CT$_{min}$ and L80 differed because we were not able to collect all data from all individuals.

displacement-by-time data, which we smoothed using quintic splines to reduce digitization error (Walker, 1998) in the R package *fda* v.5.1.4 (Ramsay et al., 2009; https://CRAN.R-project.org/package=fda). We calculated velocity curves as the first derivative of the displacement curve (Walker, 1998; Moen et al., 2013). From the velocity curves we calculated maximum velocity values at each temperature for each individual. We digitized multiple videos for each individual and temperature, and we selected the video with the highest velocity to represent peak performance for a given individual at that temperature. In total, we digitized, smoothed and analyzed videos from 781 jumps across 75 individuals. Finally, for each individual and temperature, we standardized peak values to the peak performance for that individual (across temperatures) to use in further analyses (John-Alder et al., 1988, 1989; Herrel and Bonneaud, 2012). This standardization reduces potential differences in absolute performance within and across species owing to body size (Bulté and Blouin-Demers, 2006). See Supplementary Materials and Methods for further justification.

To characterize TPCs for each species, we compared two regression models. In these models, we regressed relative peak performance on the measured temperature at which frogs achieved the performance. For each species, we compared the statistical support for quadratic and Gaussian regression functions to estimate a single curve across all individuals within a species (Table S1; Angilletta, 2006). Both models allow the typical hump-shaped form of performance curves. However, the Gaussian approach can fit TPCs better than standard quadratic regression (Angilletta, 2006). For each species, we compared the models with AICc and their associated weights (Burnham and Anderson, 2002). We performed all curve-fitting analyses with base functions in R v.4.0.2 (https://www.r-project.org/). Comparing these two models is one of many approaches proposed for characterizing TPCs (Angilletta, 2006; Bulté and Blouin-Demers, 2006; Rezende and Bozinovic, 2019), which we discuss in more detail in the Supplementary Materials and Methods.

We chose the optimal regression model for each species based on the lowest AICc. This model was used to determine the range of temperatures at which each species exhibited high performance. We used the lowest temperature at which velocity dropped to 80% of peak performance (L80; John-Alder et al., 1988) as a measure of the lowest temperature at which a species still has high jumping performance, following previous work (e.g. Huey and Stevenson, 1979; John-Alder et al., 1988, 1989; Navas, 1996; Angilletta et al., 2002; Herrel and Bonneaud, 2012; Logan, et al., 2014; Kellermann et al., 2019). We also tested alternative performance thresholds (L70, L90 and temperature of peak performance) to examine sensitivity of our results to our 80% threshold. Overall, we focused on the lower end of the TPC instead of breadth of the curve, to emphasize colonization of the (cooler) temperate zone. Moreover, previous studies of both thermal tolerances and performance curves in terrestrial ectotherms have shown that upper (i.e. hot) bounds vary little with latitude (Snyder and Weathers, 1975; Addo-Badiako et al., 2000; Sunday et al., 2011, 2019) or when comparing temperate and tropical species (John-Alder et al., 1988; van Berkum, 1988). Nonetheless, we also tested curve breadth, finding nearly identical results in phylogenetic comparative analyses (Supplementary Materials and Methods).

Cold tolerance

To test whether tolerance to cold temperatures has limited tropical hylids from colonizing the temperate zone, we estimated each species mean of the critical thermal minimum (CT$_{min}$): the lowest temperature at which an organism loses the ability to right itself (Lutterschmidt and Hutchison, 1997). When compared with alternative measures of tolerance (e.g. lower-lethal limit), CT$_{min}$ produces very similar results in interspecific comparative studies (Sunday et al., 2011, 2019).

As CT$_{min}$ trials occurred after performance trials, frogs had been acclimated to 20°C for 2 weeks before measuring CT$_{min}$. We first placed frogs in a water-filled beaker, with the water level and beaker size adjusted to submerge each frog (John-Alder et al., 1988; Muñoz et al., 2014). We then placed the beaker in an ice bath, which reduced each frog's body temperature at a constant rate of approximately 1°C min^{-1} (John-Alder et al., 1988; von May et al., 2017, 2019). This rate should produce negligible lag between changes in water temperature and the core body temperature of small amphibians (e.g. Near et al., 1990). We also used an infrared thermometer on a subset of individuals to verify concordance of body and water temperature when conducting trials, which was true across body sizes. Once each frog started to reduce movement within the water, we conducted righting-response trials every 1°C. We removed frogs from the water and stimulated them to move by touching their legs with a thin piece of plastic to keep them from remaining motionless (e.g. as a defensive tactic).

To reduce measurement error, we calculated CT$_{min}$ as the average of two temperatures: (1) the temperature at which a frog lost its ability to right itself within 60 s (Layne and Romano, 1985; John-Alder et al., 1988); and (2) the immediately preceding recorded temperature at which it could right itself. The actual CT$_{min}$ for a given individual will lie between these two temperatures. We collected CT$_{min}$ data from all species except for *T. smithii*, given an unexpected death after performance trials. We used mean±s.e.m. CT$_{min}$ for each species in phylogenetic comparative analyses.

Phylogenetic comparative analyses

We extracted a tree of our 12 study species from the amphibian phylogeny of Jetz and Pyron (2018). That study represents the most comprehensive estimate of anuran phylogeny available. It also has branch lengths in units of time, which are the most appropriate units for comparative analyses (Butler and King, 2004; O'Meara et al., 2006). Jetz and Pyron (2018) presented a Bayesian posterior distribution of time-calibrated trees, from which we calculated a consensus for our 12 species. We first used the VertLife website (www.vertlife.org/phylosubsets) to download 1000 trees from the posterior distribution, each pruned to include only our 12 taxa. We then used TreeAnnotator (Bouckaert et al., 2019) to calculate a single maximum clade credibility (MCC) tree, with branch lengths estimated as their mean values across the posterior distribution of trees (Fig. 2). All trees in this posterior distribution had identical topologies for our 12 taxa (i.e. posterior probability of 1.0 for all nodes) and showed low variation in branch lengths, suggesting that phylogenetic uncertainty would not influence our results. We thus used the resulting MCC for all analyses (see Appendix S3 in Moen et al., 2021a).

We compared four models of phenotypic evolution to test whether each physiological trait (CT$_{min}$, L80) adapted with the colonization of the temperate zone from the tropics. We conducted separate analyses on the two traits to assess them individually. We compared BM and three different OU models. Two of our OU models explicitly modeled change in selective environments, or regimes (Hansen, 1997; Butler and King, 2004), such as changing from the tropics to the temperate zone. A strong fit to a BM model would indicate that similarity among species is best described by the amount of their shared evolutionary history. Support for our second
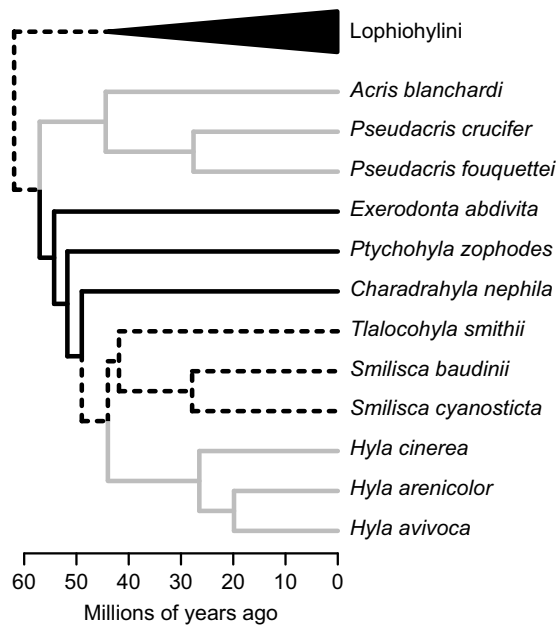
Fig. 2. Phylogeny of Middle American treefrogs in this study. Topology and branch lengths come from a maximum clade credibility consensus tree of the study taxa pruned from Jetz and Pyron (2018); all branches are supported by a posterior probability of 1.0. Temperate lineages are in gray and tropical lineages are in black. Tropical lowland and highland lineages are represented by dashed and solid branches, respectively. The Middle-American clade's sister group, Lophiohylini, is shown for comparison, with the triangle representing a larger, multi-species clade. See Fig. S1 to show how our 12 taxa fit within a more broadly sampled tree.

model, a single-optimum OU process (OU1), would indicate that species are adapting to a single selective regime for the whole clade. For our third model, we fit a two-regime OU model (OU2) in which one selective regime was for temperate species and one for tropical. Support for this third model would favor the hypothesis that CT_{min} or L80 changed the two times that hylid clades colonized the temperate zone (Fig. 2).

Tropical species of the Middle American clade can be further subdivided into species that inhabit highland and lowland tropical climates (Smith et al., 2007). Given distinct seasonal climatic zonation along elevational gradients in the tropics (Janzen, 1967; Polato et al., 2018), one might expect that each climate has a separate optimum phenotype (Navas et al., 2008). Thus, our fourth model (OU3) had three OU regimes: temperate, tropical highlands and tropical lowlands. We consider only one temperate regime because five of our six temperate species occur only at low elevations, and we collected the sixth species (*Hyla arenicolor*) at low elevations. We categorized each tropical species as highland or lowland based on elevational distributions reported in Duellman (2001) and AmphibiaWeb (2021). Species with an elevational midpoint of 500 m and below were considered as lowland, whereas species with an elevational midpoint of 1000 m and above were considered highland. Importantly, we collected all highland species in this study above 1000 m and all lowland species below 500 m.

To compare models of phenotypic evolution of CT_{min} and L80, we used the R package *OUwie* v.2.6 (Beaulieu et al., 2012; https://CRAN.R-project.org/package=OUwie) to calculate likelihood support and parameter values. We then compared models based on AICc and their weights (Burnham and Anderson, 2002). Models 2–4 only differed in OU regimes and thus optima. While optima can be estimated well at small sample sizes (Beaulieu et al., 2012; Ho

and Ané, 2013, 2014a; Cressler et al., 2015), estimating α – particularly different α for each regime – is much more challenging (Beaulieu et al., 2012; Ho and Ané, 2013). Therefore, we held α and σ^2 constant across regimes. We assumed the stationary distribution for the ancestral phenotypic state in OU models (Ho and Ané, 2014a), given that this assumption produced more reasonable optima values in OU2 and OU3. Models OU2 and OU3 required assigning internal node states for area (tropical and temperate) and elevation (lowland and highland), respectively. We assigned these states based on previous estimates for the entire Middle American clade (Smith et al., 2007; Moen et al., 2009; Pyron, 2014), which we map on the phylogeny in Fig. 2 and describe in detail in the Supplementary Materials and Methods.

For CT_{min}, we analyzed species means with their standard errors (s.e.m.) to account for intraspecific variation and measurement error (Hansen and Bartoszek, 2012; Silvestro et al., 2015). We calculated standard errors using the small sample size method of Ives et al. (2007), which involves calculating a pooled estimate of sampling variance and allows assigning standard errors to species with only a single individual (see Ives et al., 2007, their Appendix S3). We could not calculate standard errors for L80 because species values were derived from a curve fit through data of all individuals. To test the potential consequences of having no standard errors for L80, we compared model fits with and without standard errors for CT_{min} (Supplementary Materials and Methods). As a further test of the robustness of our results, we repeated analyses of L80 after excluding data from *Tlalocohyla smithii*, given that its TPC was based on a single individual and thus more likely to be affected by measurement error.

RESULTS
General patterns in thermal performance curves
We found that the majority of the species-level performance curves were unimodal curves (Table S1), with performance peaking at intermediate to high temperatures (20–35°C) and decreasing towards extreme temperatures (Figs 3 and 4). However, two of the 12 performance curves were best fit by parameters that produced a monotonically rising curve (Table S1, Figs 3 and 4), showing the highest performance at the highest temperatures.

In the temperate zone, the lowest temperatures at which velocity dropped to 80% of peak performance (L80) ranged from 10.23 to 12.28°C for early-spring breeders (*Acris blanchardi*, *Pseudacris crucifer* and *P. fouquettei*). Late-spring breeders (*Hyla cinerea*, *H. avivoca* and *H. arenicolor*) showed higher L80 values, ranging from 14.78 to 16.46°C (Table 1, Fig. 3). L80 values for tropical highland and lowland species overlapped, with species means ranging from 10.67 to 17.59°C for highland species and from 13.74 to 27.45°C for lowland species (Table 1, Fig. 4).

General patterns in thermal tolerance (CT_{min})
In the temperate zone, CT_{min} was lower in early-spring breeders (0.27–0.56°C) than in late-spring breeders (2.86–5.44°C; Table 1). Likewise, CT_{min} among highland tropical species (5.14–6.42°C) did not overlap with that of lowland species (6.90–8.92°C; Table 1). However, differences in CT_{min} in tropical lowland and highland species were generally much smaller than between early- and late-breeding temperate species (Table 1). Moreover, nearly all tropical species showed higher CT_{min} values than temperate species (Table 1).

Evolution of L80 and CT_{min}
BM was the most highly supported model for L80, with an AICc weight of 0.778 (Table 2). This means that similarity among species
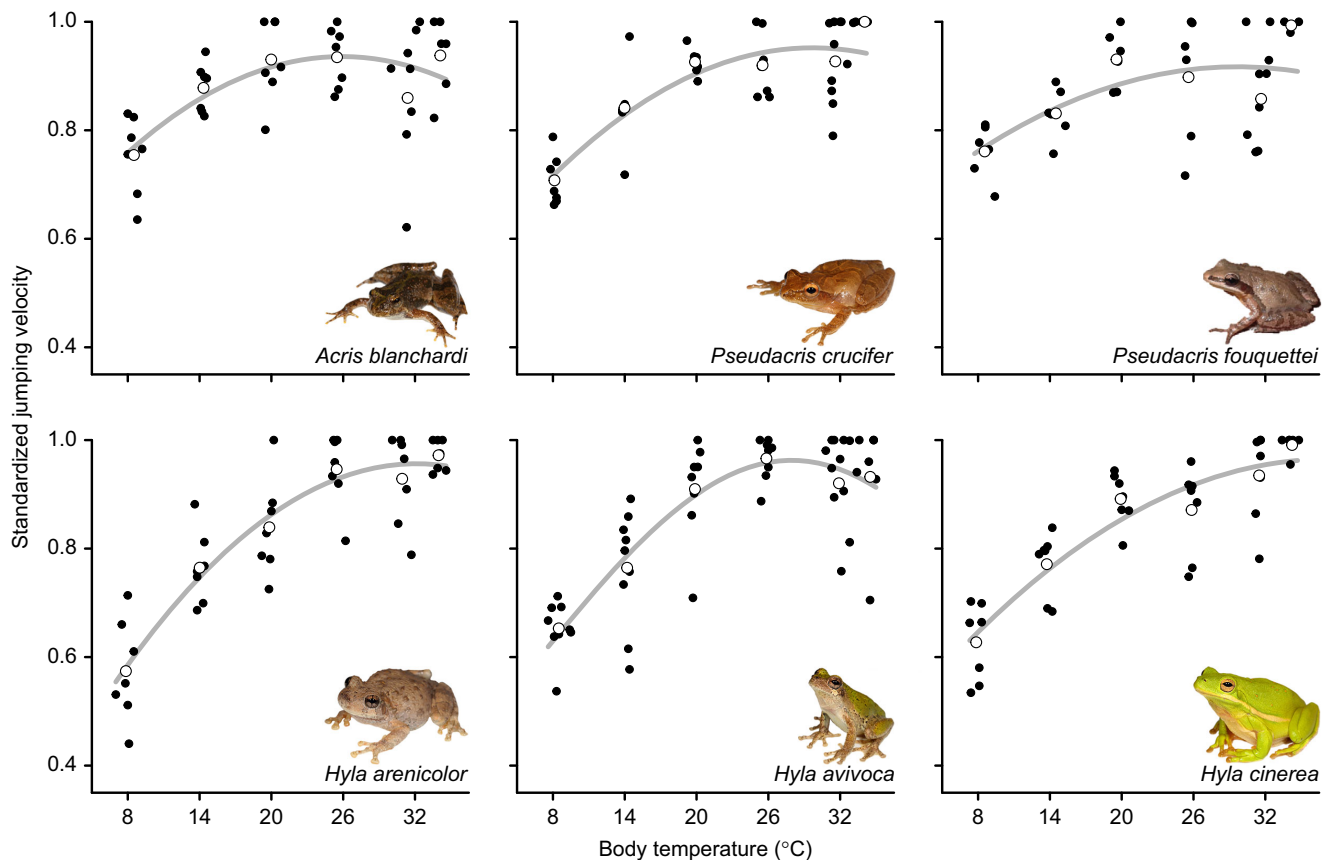
Fig. 3. Thermal performance curves (TPCs) used to estimate the lower temperature threshold for 80% of peak jumping performance (L80) of six temperate species. Species in the top row are early-spring breeders, whereas those in the bottom row are late-spring breeders. Gray lines represent TPCs, with the functional form of the curve following the optimal model for each species (Table S1). Black dots represent individual performance. White dots represent mean performance for species at a given temperature. Note that because jumping performance is standardized to peak performance within individuals, this variable is unitless. Photos by A.R.H. (*Pseudacris fouquettei*) and D.S.M. (all others).

was best described by shared evolutionary history, rather than biogeographic region. The only other model with appreciable support was the single-optimum OU1 model (AICc weight=0.179), which means species are adapting to a single phenotypic optimum for the whole clade. We found nearly identical results for L70, L90, temperature of peak performance, and the lower half of thermal performance breadth (Table S2). Furthermore, our results were the same with and without *Tlalocohyla smithii* (Table S3), the species whose TPC we estimated from just one individual (Fig. 4). Thus, our results were robust to both the method of characterizing jumping performance at low temperatures, as well as to inclusion of a species whose TPC was based on few data.

We found that CT_{min} decreased when hylids colonized the temperate zone. The model with the most support for CT_{min} evolution was OU2 (AICc weight=0.729; Table 2), which had one CT_{min} optimum for temperate species (−1.40°C) and another, much higher, for tropical species (6.50°C; Table 3, Fig. 5). Results were nearly identical when excluding standard errors (Table S4), suggesting that L80 analyses were likewise robust to exclusion of standard errors.

DISCUSSION
In this study, we tested how thermal tolerances and performance curves evolved with the colonization of the temperate zone from the tropics in six tropical and six temperate hylid treefrog species. We found that CT_{min} most likely adapted by decreasing upon colonization of temperate climates. However, L80 showed no

adaptation associated with biogeographic region. Our results suggest that increasing tolerance to low temperatures was a key to colonizing the temperate zone, whereas improving jumping performance at low temperatures was not. Here, we discuss these results further and address the prospects and challenges of implementing the OU process in studies of physiological adaptation.

Physiology of colonizing the temperate zone
We found that colonization of the temperate zone led to an adaptive reduction in CT_{min} in Middle American treefrogs. In a previous study on the latitudinal gradient of species diversity in hylid frogs, Wiens et al. (2006) examined the northern range limits of four hylid clades that have colonized tropical Middle America from South America, but have not reached the temperate zone. They showed that these limits were associated with higher temperature seasonality and lower annual temperature extremes. Our results suggest that the ecophysiological explanation for this pattern is niche conservatism in CT_{min} in tropical hylid frogs, limiting colonization of the temperate zone from the tropics.

Despite niche conservatism in most hylid clades, Middle American hylids colonized temperate North America twice from the lowland tropics (Smith et al., 2005; Moen et al., 2009). The *Acris–Pseudacris* clade arrived first, and species in this clade are generally early-spring breeders. For example, *Pseudacris crucifer* becomes active during winter months and can tolerate freezing temperatures (Schmid, 1982; Dodd, 2013), which is consistent with our CT_{min} results for species in this clade (Table 1). In contrast, *Hyla*
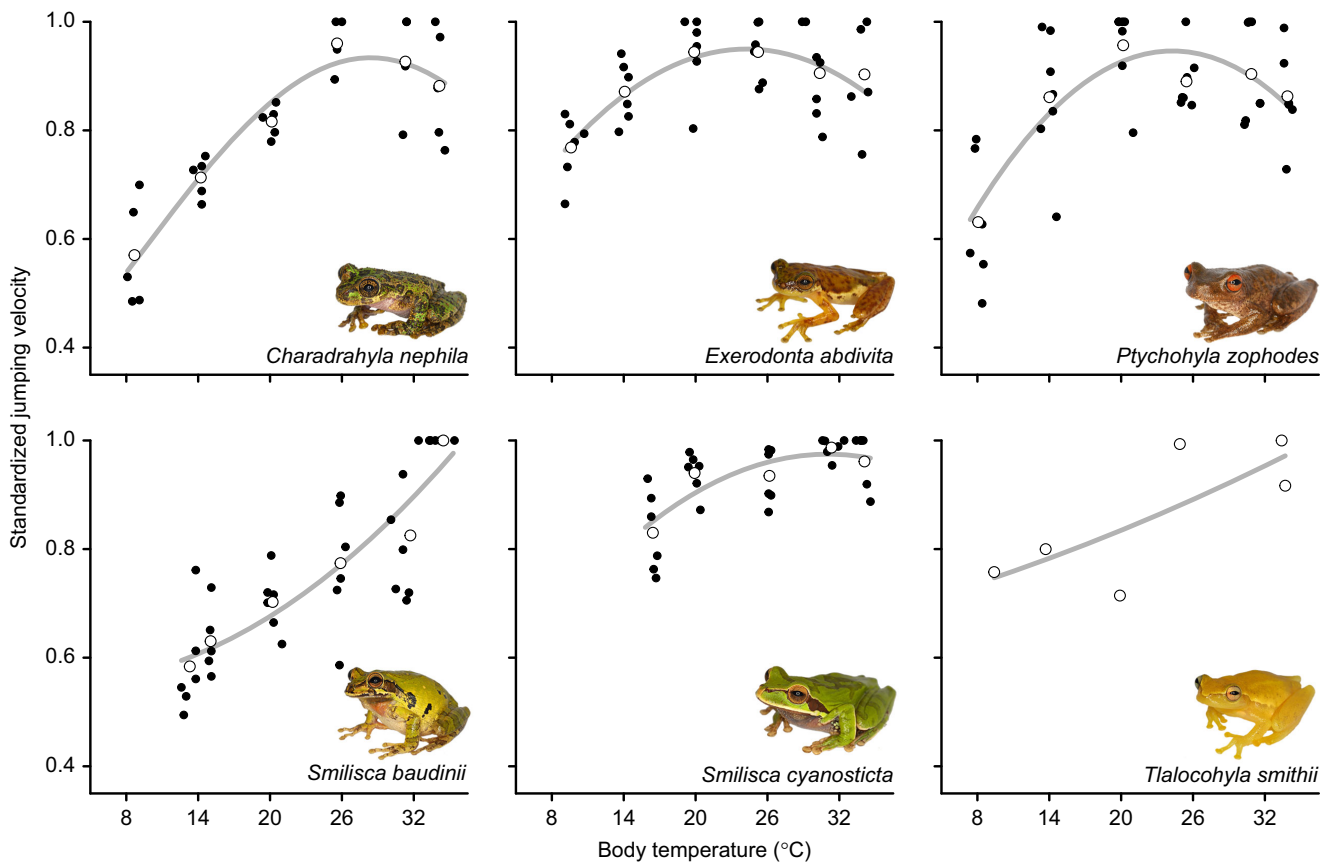
Fig. 4. TPCs used to estimate the L80 of six tropical species. Figure details follow those of Fig. 3. Highland species are in the top row; lowland species are in the bottom row. All photos taken by D.S.M.

colonized temperate North America later and comprises late-spring breeders (Dodd, 2013). Species in this clade have less cold tolerance than members of the *Acris–Pseudacris* clade, but more tolerance than tropical species (Table 1). However, how freezing tolerance relates to CT_{min} across species remains unclear. In particular, *Hyla versicolor* and *H. chrysocelis* can survive freezing temperatures (Schmid, 1982; Costanzo et al., 1992; Storey and Storey, 1992),

despite being in the temperate clade with higher CT_{min} than the *Acris–Pseudacris* clade. Whether other *Hyla* species can survive freezing is an intriguing direction for future work.

In contrast to CT_{min}, biogeographic region did not affect L80 evolution. Our results resemble those of MacLean et al. (2019), who found that cold tolerance was associated with latitude in *Drosophila*, but aspects of TPCs (e.g. breadth, optimal temperature) were not. A similar contrast between tolerances and performance has also been found in actively thermoregulating taxa (van Berkum, 1988). Though frogs generally do not thermoregulate (Navas, 1996; Navas et al., 2008), they may use seasonal activity as a long-term thermoregulation strategy, waiting for optimal temperatures before becoming active (e.g. to reproduce, locate prey, disperse). But without migration, individuals must tolerate all climatic conditions

Table 2. Models of physiological trait evolution

Variable: model	k	ln L	AICc	w_i
Tolerance (CT_{min})				
Brownian motion	2	−24.720	54.941	0.209
OU single optimum	3	−25.870	61.170	0.009
OU2	**4**	**−18.888**	**52.442**	**0.729**
OU3	5	−17.843	57.686	0.053
Jumping performance (L80)				
Brownian motion	**2**	**−35.342**	**76.018**	**0.778**
OU single optimum	3	−34.977	78.953	0.179
OU2	4	−34.220	82.154	0.036
OU3	5	−32.838	85.677	0.006

We compared four models, including Brownian motion and three Ornstein–Uhlenbeck (OU) models. OU single optimum: one optimal temperature for all species. OU2: separate temperature optima for tropical and temperate lineages. OU3: separate temperature optima for temperate, lowland tropical and highland tropical lineages. k, number of free parameters in the model; ln L, log-likelihood; AICc, corrected Akaike information criterion; w_i, AICc model weight. Note that Brownian motion is not nested within OU models in the way we estimated them (O'Meara and Beaulieu, 2014) and so may not necessarily have a lower log-likelihood, as for CT_{min}. For each variable, we indicate the optimal model (i.e. the lowest AICc and highest w_i) in bold.

Table 3. Parameter estimates for optimal models of CT_{min} and L80 evolution

Variable	CT_{min}	L80
Optimal model	OU2	Brownian motion
σ^2	3.378±0.843	24.963±0.408
α	1.063±0.880	–
Ancestral state	–	13.959±2.051
Temperate θ	−1.401±1.362	–
Tropical θ	6.504±0.739	–

σ^2 is the rate of stochastic evolution, with units of °C². Ancestral state is in °C and was not estimated for OU models (Ho and Ané, 2014a). In OU models, θ is the primary optimum in °C for each climate regime and α is the strength of pull towards optima, with units of millions of years^{-1}. All values represent maximum-likelihood estimates with one standard error.
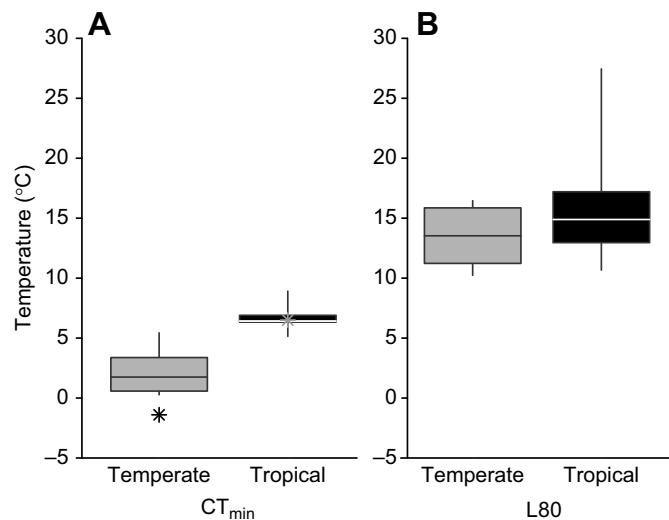
Fig. 5. Boxplot showing the distribution of critical thermal minimum (CT$_{min}$) and L80 for tropical and temperate study species means (Table 1). (A) CT$_{min}$ and (B) L80. Lines in the center of boxes are medians. Lower and upper edges of boxes are the 25 and 75% quartiles, respectively. Whiskers represent the most extreme species means for each group. Asterisks for CT$_{min}$ indicate the OU optima for each region's taxa, based on the OU2 model.

during the whole year, even if inactive during much of that time (Ludwig et al., 2015). Thus, selection may more strongly favor tolerance to low temperatures upon colonizing the temperate zone than the ability to function at colder temperatures.

Given that *Hyla* and members of the *Acris–Pseudacris* clade show largely overlapping distributions (Duellman and Sweet, 1999), why do they differ in CT$_{min}$ and L80 (Table 1)? The answer might relate to time since colonization and evolutionary 'starting point'. The majority of Middle American hylids are found in the tropics at high elevations (Duellman, 2001), with lowland species nested within largely highland clades (Smith et al., 2007). Thus, species with physiology fitting highland climates may have made the initial colonization of Middle America, then later expanded to lowlands (Wiens et al., 2006; Moen et al., 2009). The *Hyla* clade that most recently colonized temperate North America (Fig. 2) shares its most recent common ancestor with the *Smilisca–Tlalocohyla* clade, the clade with the highest CT$_{min}$. This could explain why *Hyla* generally showed higher CT$_{min}$ and L80 than members of the *Acris–Pseudacris* clade: their evolutionary starting point was higher, and they have been adapting to a temperate climate for less time. We note that this 'time for adaptation' forms an integral part of the OU model of adaptation (Fig. 1; Butler and King, 2004; Moen et al., 2016). Other approaches (e.g. phylogenetic ANOVA) assume all lineages have been in their current environment for their entire history (Hansen and Orzack, 2005; Hansen, 2014), potentially compromising the study of adaptation.

Lastly, we acknowledge that future work should consider additional species and a higher maximum temperature for characterizing TPCs. While our results clearly favored different CT$_{min}$ optima for temperate and tropical hylids, the optimum for temperate taxa was below freezing (−1.40°C). This value is warmer than lethal thermal minima for many temperate anurans (Brattstrom, 1968), and thus reasonable, yet it was lower than the CT$_{min}$ of any species we measured here (Table 1). Moreover, additional simulations suggested that increasing species sampling may increase our ability to distinguish OU3 from OU2 and allow for

more precise estimates of OU optima (see Supplementary Materials and Methods, Fig. S2). We also found monotonically rising TPCs in two tropical lowland species (Fig. 4), which means the temperature of their peak performance may be higher than our highest tested temperature (35°C). If true, their L80 values here would be underestimates. Future work should thus consider higher temperatures, particularly for tropical lowland species.

Remaining questions in thermal physiology and biogeography

Both our approach and our results raise several questions for future research on the physiology underlying colonization of the temperate zone. First, are the relevant factors in hylids the same as those important in other organisms, including other families of anurans? Members of two additional families have also colonized temperate North America from the Neotropics: Bufonidae (Mendelson et al., 2011; Pyron, 2014; Portik and Papenfuss, 2015) and Microhylidae (Streicher et al., 2012; Pyron, 2014). These families differ from hylids in ecology and climatic niche (Dodd, 2013; Moen and Wiens, 2017). Thus, studying them could address whether tolerances matter more than locomotor performance (present study; MacLean et al., 2019), or if instead our results are specific to particular aspects of ecology. Moreover, examining additional taxa could determine the importance of time of colonization, as hypothesized above for *Hyla*.

Second, what role do other aspects of physiology play in colonizing the temperate zone from the tropics? Metabolism has often been compared in temperate and tropical amphibians. These studies have shown that low-temperature metabolism is higher in temperate than in tropical taxa (Brattstrom, 1968; Hutchison et al., 1968; Whitford, 1973; Snyder and Weathers, 1975; Feder, 1978, 1982; Walton, 1993; Navas, 1996). Future evolutionary studies that test adaptation in metabolism may reveal an additional important factor in determining patterns of species richness.

Finally, what role does acclimation play in colonization of the temperate zone? In this study we accounted for acclimation by holding individuals at a constant temperature for 1 week (TPCs) or 2 weeks (CT$_{min}$) prior to data collection, following previous studies. Moreover, most studies show no acclimation to cold temperatures in locomotor performance in adult anurans (Putnam and Bennett, 1981; Whitehead et al., 1989; Knowles and Weigl, 1990; Wilson and Franklin, 2000; but see Padilla et al., 2019), particularly in non-aquatic species (Wilson et al., 2000), which we studied here. However, nearly all anuran acclimation studies have been performed on temperate species. A synthesis across ectothermic taxa suggested that tropical organisms may show a higher capacity for physiological acclimation (Seebacher et al., 2015), but comparisons of temperate and tropical amphibians show the opposite pattern in metabolism (Feder, 1978, 1982). Moreover, acclimation effects on CT$_{max}$ in anurans seem much weaker in tropical species (Brattstrom, 1968; Mahoney and Hutchison, 1969; Christian et al., 1988; Navas et al., 2008), and no work of which we are aware has examined how acclimation affects CT$_{min}$ in anurans. If acclimation has been important in past colonization of the temperate zone in hylids, we would expect that lineages closely related to successful colonists (e.g. *Smilisca* closely related to *Hyla*; Fig. 2) would show high propensity to reduce CT$_{min}$ after acclimation to lower temperatures. For example, *Smilisca baudinii* has a broad climatic distribution, from the tropical lowlands (where we sampled) to populations in south Texas, USA. Understanding whether this broad distribution is due to genetic change, plasticity or both could clarify the role of acclimation in colonization of temperate climates.

Prospects and challenges of the Hansen model in comparative physiology

The Hansen model has seen limited use in comparative physiology. Part of the reason is that analyzing the model is less direct than analogous approaches, such as phylogenetic ANOVA. Uncertainties in best practices (Ho and Ané, 2014a; O'Meara and Beaulieu, 2014) lead to further confusion, and statistical properties of the method have been questioned in recent years (e.g. Boettiger et al., 2012; Ho and Ané, 2014a). Thus, we describe here six key hurdles to using the Hansen model, and we outline strategies to explicitly address these issues in an empirical study. We also provide R Markdown tutorials to help readers understand how to use the model and interpret results (Moen et al., 2021a, Appendices S10, S14). Our discussion reveals that many problems are less severe than often characterized in the literature, and we remain optimistic about more widespread use of the model to test adaptation.

(1) Ancestral-state estimation

A key implementation challenge of the Hansen model with discrete environments is specifying environments at internal nodes. Here, we sampled few species from our clade of interest (12 of 197 species; AmphibiaWeb, 2021). Thus, we used previously estimated states based on a much larger sample (Moen et al., 2009) to prevent inaccuracy owing to insufficient sampling (Salisbury and Kim, 2001). However, estimation of internal states for OU models is not trivial (e.g. Moen et al., 2016), and such estimates have been criticized for their uncertainty or inaccuracy (Cunningham et al., 1998; Cunningham, 1999; Losos, 1999). An increasingly adopted solution to overcome these challenges is using Bayesian stochastic character mapping to generate many possible character histories (Huelsenbeck et al., 2003; Bollback, 2006). One can then fit the Hansen model on each character history and integrate results over the uncertainty of internal states (e.g. Price et al., 2015; Grossnickle et al., 2020; Corn et al., 2021). Alternatively, a likelihood-based approach can be used to account for internal-state uncertainty, as outlined by Cressler et al. (2015). However, implementation of this approach is less straightforward than stochastic character mapping. Regardless of the strategy chosen for internal-state estimates, the exponential decay in the influence of adaptation to past environments means that a strong adaptive pull (i.e. high α) will quickly erase the influence of internal node states (Hansen, 1997; Butler and King, 2004). Therefore, when α is high, OU results are likely robust to ancestral-state uncertainty (Moen et al., 2016).

(2) Different statistical designs

Some types of predictor variables (e.g. continuous) and more complicated statistical designs (e.g. ANCOVA, multivariate) will present their own challenges. In particular, multivariate OU models (i.e. multiple traits adapting to the same environments) can require estimation of additional parameters (Bartoszek et al., 2012). It is also unclear whether the implementation of the multivariate, multi-optimum model (Bartoszek et al., 2012) shares some of the same statistical problems of multivariate, single-optimum OU models (Adams and Collyer, 2018). Further model developments, such as those of Clavel and Morlon (2020) and Clavel et al. (2019), may be necessary to ensure optimal statistical performance of OU models.

(3) Data from few species

Studies in comparative physiology are often limited by number of study species, given the difficulty of data collection (Lauder, 1990, 2003; Garland et al., 2005). For example, we collected and digitized high-speed video of 781 jumps at five field locations across two countries to obtain peak jumping performance of 75 individuals at six temperatures. Yet this effort rendered L80 values for just 12 species for our phylogenetic comparative analyses. Given that the simplest version of the Hansen model requires estimating more parameters than equivalent PGLS or ANOVA tests, researchers have right to worry when they have data for few species (Cooper et al., 2016a). However, the effect of sampling is more nuanced than simply comparing the number of parameters to species. For example, α may be hard to estimate regardless of species number (Beaulieu et al., 2012; Ho and Ané, 2013, 2014a; Cressler et al., 2015), but the optima can be estimated well with few species (Ho and Ané, 2013; Cressler et al., 2015), particularly when at least one regime convergently originates two or more times in the phylogeny (Ho and Ané, 2014a). Simulations have shown that effect size, rather than sample size per se, may matter more for estimating optima (Ho and Ané, 2013, 2014a) and for comparing models that only differ in number of optima (Boettiger et al., 2012; Cressler et al., 2015). Our empirical results were consistent with these simulation results: with just 12 species, we found strong evidence for an OU model with multiple optima in CT_{min} (Table 2) because these optima were distinct (Table 3, Fig. 5). Moreover, parametric bootstrapping showed that comparison of our top OU model (OU2) with the simpler model with the highest support (BM) was statistically robust (Fig. S2), further suggesting that concern about small sample sizes really depends on the particular dataset. More generally, because optima are often the focus of model comparison (e.g. present study; Scales et al., 2009; Collar et al., 2011; Grossnickle et al., 2020), these results are promising for comparative physiology.

(4) Problems with model selection and parameter estimation

One criticism of model selection in OU studies is that even modest measurement error (if ignored in the analysis) can bias AIC-based model comparison to favor a single-optimum OU model over BM (Cooper et al., 2016b). To avoid this problem, most OU implementations accommodate measurement error (Beaulieu et al., 2012; Hansen and Bartoszek, 2012; Ho and Ané, 2014b). Moreover, the multi-optimum Hansen model shows much less error in model selection than comparing single-optimum OU models to BM (Cressler et al., 2015). A second criticism is that AIC-based model selection can be biased to favor more complex OU regime mappings, elevating Type-I error rates (Boettiger et al., 2012). As we have shown (Fig. S2), this is an empirical problem that depends on each individual dataset. Many of our own empirical studies have shown highest statistical support for models of intermediate, rather than highest, complexity (present study; Moen et al., 2016; Moen, 2019). Nonetheless, researchers can conduct parametric bootstrapping analyses to examine how much information exists in their data to distinguish models (Boettiger et al., 2012). A third statistical criticism is that parameter estimates can be imprecise, even with many species. α and σ^2 can be hard to estimate separately, as high values of both parameters may be as likely as low values of both (O'Meara and Beaulieu, 2014). One solution is to focus on the estimate of stationary variance, which is a ratio of the two ($\sigma^2/2\alpha$), rather than the two parameters separately (e.g. Price et al., 2015). Estimating multiple α values is challenging for even very large datasets, so α may be best left constant across regimes when analyzing small datasets (Beaulieu et al., 2012; O'Meara and Beaulieu, 2014). Finally, when selective regimes originate only a single time on a tree, different optima can be impossible to distinguish (Ho and Ané, 2014a), so it is important to ensure at least one regime occurs on different parts of the tree. In summary, while

researchers undoubtedly need to be aware of these concerns, many of them can be addressed with careful analysis.

(5) Testing the robustness of results

Currently, several tools are available to directly test the robustness of the results obtained from the Hansen model. First, the R package *OUwie* allows users to conduct eigenanalysis of the likelihood surface for α and σ^2 to verify whether their estimates are stable. Second, both packages *ouch* and *OUwie* allow parametric bootstrapping of optimal models to calculate 95% confidence intervals for parameters (Scales et al., 2009; Boettiger et al., 2012). Third, parametric simulations to test statistical power and error rate are conveniently implemented in the R package *pmc* (Boettiger et al., 2012), which uses *ouch* to fit OU models. Such simulations are less developed to work with *OUwie*, so we have written a function that works similar to pmc but simulates data and analyzes models with *OUwie*. We include this function in Appendix S5 (Moen et al., 2021a) and demonstrate its use in the Supplementary Materials and Methods, where we examine the statistical properties of our dataset with parametric bootstrapping. We also recognize that methods are more frequently used when more resources exist to guide researchers. Thus, we demonstrate with tutorials how to use the three strategies to test for robustness of results (Moen et al., 2021a, Appendices S10, S14). Moreover, we encourage collaboration between researchers collecting empirical data and those more experienced in phylogenetic comparative methods. Such collaboration will lead to development and testing of more robust analyses and additional biological insight (Lauder, 2003; Cooper et al., 2016a; Waldrop and Rader, 2020).

(6) Potential misuse of the OU process

We caution against using the OU process in standard PGLS regression (Cooper et al., 2016b), as such analyses model evolution very differently than described in this paper (Hansen, 2014; Pennell, 2015). Instead of modeling adaptation of phenotypes to adaptive optima, typical PGLS analyses use the OU process to model residual variation around a regression line. This use has the same problems that we described in the Introduction for BM: inherited maladaptation and inability to simultaneously estimate the mean and error structure of a model (Hansen and Orzack, 2005; Hansen, 2014). When testing adaptation of a trait to a continuous environment, we encourage researchers to instead use the approach described by Hansen et al. (2008), which models adaptation as described throughout this paper.

Conclusions

Here, we reviewed the Hansen model to increase its accessibility and application in comparative physiology and biomechanics. In our case study, we examined the thermal physiology associated with biogeographic dispersal from the tropics to the temperate zone in Middle American treefrogs. We found cold tolerance (CT_{min}) important for explaining the transition, whereas jumping well at low temperatures (L80) was not. Our results suggest that the niche conservatism that has prevented most hylid clades from colonizing the temperature zone relates to intolerance to its seasonally cold temperatures. However, the physiological basis for such conservatism in other ectotherms has not been explicitly tested with phylogenetic analysis. Future work should use OU models of adaptation to examine the importance of cold tolerance, as compared with other physiological factors such as locomotion. Testing more factors, such as metabolism and acclimation, will also enrich our understanding of the origins of species richness patterns

that vary over climatic gradients. As data are increasingly available for large-scale comparative studies in physiology and biomechanics (Higham et al., 2021), we anticipate that OU models will aid in testing the role of different traits in driving patterns of physiological and species diversity.

Data availability

All supplementary data are available from the Dryad Digital Repository (Moen et al., 2021a): https://doi.org/10.5061/dryad.t4b8gtj2m. R code and R Markdown tutorials are hosted on Zenodo (https://doi.org/10.5281/zenodo.5165563 and https://doi.org/10.5281/zenodo.5165565) and also available via a link on Dryad.

References

Adams, D. C. and Collyer, M. L. (2018). Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations. *Syst. Biol.* **67**, 14-31. doi:10.1093/sysbio/syx055

Addo-Bediako, A., Chown, S. L. and Gaston, K. J. (2000). Thermal tolerance, climatic variability and latitude. *Proc. R. Soc. B* **267**, 739-745. doi:10.1098/rspb.2000.1065

AmphibiaWeb (2021). AmphibiaWeb: Information on amphibian biology and conservation. Accessed 28 September 2021. http://amphibiaweb.org.

Amundson, R. (1996). Historical development of the concept of adaptation. In *Adaptation* (ed. M. R. Rose and G. V. Lauder), pp. 11-53. San Diego: Academic Press.

Angilletta, M. J., Jr. (2006). Estimating and comparing thermal performance curves. *J. Therm. Biol.* **31**, 541-545. doi:10.1016/j.jtherbio.2006.06.002

Angilletta, M. J., Jr, Hill, T. and Robson, M. A. (2002). Is physiological performance optimized by thermoregulatory behavior?: a case study of the eastern fence lizard, *Sceloporus undulatus*. *J. Therm. Biol.* **27**, 199-204. doi:10.1016/S0306-4565(01)00084-5

Araújo, M. B., Ferri-Yáñez, F., Bozinovic, F., Marquet, P. A., Valladares, F. and Chown, S. L. (2013). Heat freezes niche evolution. *Ecol. Lett.* **16**, 1206-1219. doi:10.1111/ele.12155

Bartoszek, K., Pienaar, J., Mostad, P., Andersson, S. and Hansen, T. F. (2012). A phylogenetic comparative method for studying multivariate adaptation. *J. Theor. Biol.* **314**, 204-215. doi:10.1016/j.jtbi.2012.08.005

Bauer, U., Poppinga, S. and Müller, U. K. (2020). Mechanical ecology—taking biomechanics to the field. *Integr. Comp. Biol.* **60**, 820-828. doi:10.1093/icb/icaa018

Baum, D. A. and Larson, A. (1991). Adaptation reviewed: a phylogenetic methodology for studying character macroevolution. *Syst. Zool.* **40**, 1-18. doi:10.1093/sysbio/40.1.1

Beaulieu, J. M., Jhwueng, D.-C., Boettiger, C. and O'Meara, B. C. (2012). Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution* **66**, 2369-2383. doi:10.1111/j.1558-5646.2012.01619.x

Bennett, J. M., Sunday, J., Calosi, P., Villalobos, F., Martínez, B., Molina-Venegas, R., Araújo, M. B., Algar, A. C., Clusella-Trullas, S., Hawkins, B. A. et al. (2021). The evolution of critical thermal limits of life on Earth. Nat. Commun. 12, 1198. doi:10.1038/s41467-021-21263-8

Blomberg, S. P., Garland, T. J., Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57, 717-745. doi:10.1111/j.0014-3820.2003.tb00285.x

Boettiger, C., Coop, G. and Ralph, P. (2012). Is your phylogeny informative? Measuring the power of comparative methods. Evolution 66, 2240-2251. doi:10.1111/j.1558-5646.2011.01574.x

Bollback, J. P. (2006). SIMMAP: stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics 7, 88. doi:10.1186/1471-2105-7-88

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N. et al. (2019). BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comp. Biol. 15, e1006650. doi:10.1371/journal.pcbi.1006650

Bozinovic, F., Calosi, P. and Spicer, J. I. (2011). Physiological correlates of geographic range in animals. Annu. Rev. Ecol. Evol. Syst. 42, 155-179. doi:10.1146/annurev-ecolsys-102710-145055

Brattstrom, B. H. (1968). Thermal acclimation in anuran amphibians as a function of latitude and altitude. Comp. Biochem. Physiol. 24, 93-111. doi:10.1016/0010-406x(68)90961-4

Bulté, G. and Blouin-Demers, G. (2006). Cautionary notes on the descriptive analysis of performance curves in reptiles. J. Therm. Biol. 31, 287-291. doi:10.1016/j.jtherbio.2005.11.030

Burnham, K. P. and Anderson, D. R. (2002). Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. New York: Springer-Verlag.

Butler, M. A. and King, A. A. (2004). Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am. Nat. 164, 683-695. doi:10.1086/426002

Calosi, P., Bilton, D. T., Spicer, J. I., Votier, S. C. and Atfield, A. (2010). What determines a species' geographical range? Thermal biology and latitudinal range size relationships in European diving beetles (Coleoptera: Dytiscidae). J. Anim. Ecol. 79, 194-204. doi:10.1111/j.1365-2656.2009.01611.x

Chown, S. L., Gaston, K. J. and Robinson, D. (2004). Macrophysiology: large-scale patterns in physiological traits and their ecological implications. Func. Ecol. 18, 159-167. doi:10.1111/j.0269-8463.2004.00825.x

Christian, K. A., Nunez, F., Clos, L. and Diaz, L. (1988). Thermal relations of some tropical frogs along an altitudinal gradient. Biotropica 20, 236-239. doi:10.2307/2388239

Clavel, J. and Morlon, H. (2020). Reliable phylogenetic regressions for multivariate comparative data: illustration with the MANOVA and application to the effect of diet on mandible morphology in phyllostomid bats. Syst. Biol. 69, 927-943. doi:10.1093/sysbio/syaa010

Clavel, J., Aristide, L. and Morlon, H. (2019). A penalized likelihood framework for high-dimensional phylogenetic comparative methods and an application to New-World monkeys brain evolution. Syst. Biol. 68, 93-116. doi:10.1093/sysbio/syy045

Collar, D. C., Schulte, J. A., II and Losos, J. B. (2011). Evolution of extreme body size disparity in monitor lizards (Varanus). Evolution 65, 2664-2680. doi:10.1111/j.1558-5646.2011.01335.x

Cooper, N., Thomas, G. H. and FitzJohn, R. G. (2016a). Shedding light on the 'dark side' of phylogenetic comparative methods. Methods Ecol. Evol. 7, 693-699. doi:10.1111/2041-210X.12533

Cooper, N., Thomas, G. H., Venditti, C., Meade, A. and Freckleton, R. P. (2016b). A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. Biol. J. Linn. Soc. 118, 64-77. doi:10.1111/bij.12701

Corn, K. A., Martinez, C. M., Burress, E. D. and Wainwright, P. C. (2021). A multifunction trade-off has contrasting effects on the evolution of form and function. Syst. Biol. 70, 681-693. doi:10.1093/sysbio/syaa091

Costanzo, J. P., Wright, M. F. and Lee, R. E. J. (1992). Freeze tolerance as an overwintering adaptation in Cope's grey treefrog (Hyla chrysoscelis). Copeia 1992, 565-569. doi:10.2307/1446222

Cowles, R. B. and Bogert, C. M. (1944). A preliminary study of the thermal requirements of desert reptiles. Bull. Am. Mus. Nat. Hist. 83, 265-296.

Cressler, C. E., Butler, M. A. and King, A. A. (2015). Detecting adaptive evolution in phylogenetic comparative analysis using the Ornstein-Uhlenbeck model. Syst. Biol. 64, 953-968. doi:10.1093/sysbio/syv043

Cunningham, C. W. (1999). Some limitations of ancestral character-state reconstruction when testing evolutionary hypotheses. Syst. Biol. 48, 665-674. doi:10.1080/106351599260238

Cunningham, C. W., Omland, K. E. and Oakley, T. H. (1998). Reconstructing ancestral character states: a critical reappraisal. Trends Ecol. Evol. 13, 361-366. doi:10.1016/S0169-5347(98)01382-2

Dodd, C. K. J. (2013). Frogs of the United States and Canada. Baltimore: Johns Hopkins University Press.

Duellman, W. E. (1999). Patterns of Distribution of Amphibians: A Global Perspective. Baltimore: Johns Hopkins University Press.

Duellman, W. E. (2001). The Hylid Frogs of Middle America. Lawrence: Society for the Study of Amphibians and Reptiles.

Duellman, W. E. and Sweet, S. S. (1999). Patterns of distribution of amphibians: a global perspective. In Patterns of Distribution of Amphibians: A Global Perspective (ed. W. E. Duellman), pp. 31-110. Baltimore: Johns Hopkins University Press.

Dunlap, D. G. (1980). Comparative effects of thermal acclimation and season on metabolic compensation to temperature in the hylid frogs, Pseudacris triseriata and Acris crepitans. Comp. Biochem. Physiol. A Physiol. 66, 243-249. doi:10.1016/0300-9629(80)90158-9

Escudero, M., Hipp, A. L., Hansen, T. F., Voje, K. L. and Luceño, M. (2012). Selection and inertia in the evolution of holocentric chromosomes in sedges (Carex, Cyperaceae). New Phytol. 195, 237-247. doi:10.1111/j.1469-8137.2012.04137.x

Faivovich, J., Pereyra, M. O., Luna, M. C., Hertz, A., Blotto, B. L., Vásquez-Almazán, C. R., McCranie, J. R., Sánchez, D. A., Baêta, D., Araujo-Vieira, K. et al. (2018). On the monophyly and relationships of several genera of Hylini (Anura: Hylidae: Hylinae), with comments on recent taxonomic changes in hylids. S. Am. J. Herpetol. 13, 1-32. doi:10.2994/SAJH-D-17-00115.1

Farrell, B. D., Mitter, C. and Futuyma, D. J. (1992). Diversification at the insect-plant interface. Bioscience 42, 34-42. doi:10.2307/1311626

Feder, M. E. (1978). Environmental variability and thermal acclimation in Neotropical and temperate zone salamanders. Physiol. Zool. 51, 7-16. doi:10.1086/physzool.51.1.30158660

Feder, M. E. (1982). Environmental variability and thermal acclimation of metabolism in tropical anurans. J. Therm. Biol. 7, 23-28. doi:10.1016/0306-4565(82)90015-8

Felsenstein, J. (1985). Phylogenies and the comparative method. Am. Nat. 125, 1-15. doi:10.1086/284325

Felsenstein, J. (1988). Phylogenies and quantitative characters. Annu. Rev. Ecol. Syst. 19, 445-471. doi:10.1146/annurev.es.19.110188.002305

Gans, C. and Parsons, T. S. (1966). On the origin of the jumping mechanism in frogs. Evolution 20, 92-99. doi:10.1111/j.1558-5646.1966.tb03345.x

Garamszegi, L. Z. (2014). Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology. Berlin: Springer-Verlag.

Garland, T., Jr, Bennett, A. F. and Rezende, E. L. (2005). Phylogenetic approaches in comparative physiology. J. Exp. Biol. 208, 3015-3035. doi:10.1242/jeb.01745

Gaston, K. J., Chown, S. L., Calosi, P., Bernardo, J., Bilton, D. T., Clarke, A., Clusella-Trullas, S., Ghalambor, C. K., Konarzewski, M., Peck, L. S. et al. (2009). Macrophysiology: a conceptual reunification. Am. Nat. 174, 595-612. doi:10.1086/605982

Gould, S. J. and Vrba, E. S. (1982). Exaptation—a missing term in the science of form. Paleobiology 8, 4-15. doi:10.1017/S0094837300004310

Grossnickle, D. M., Chen, M., Wauer, J. G. A., Pevsner, S. K., Weaver, L. N., Meng, Q. J., Liu, D., Zhang, Y. G. and Luo, Z. X. (2020). Incomplete convergence of gliding mammal skeletons. Evolution 74, 2662-2680. doi:10.1111/evo.14094

Gvoždík, L. and Van Damme, R. (2006). Triturus newts defy the running-swimming dilemma. Evolution 60, 2110-2121. doi:10.1554/06-070.1

Hansen, T. F. (1997). Stabilizing selection and the comparative analysis of adaptation. Evolution 51, 1341-1351. doi:10.1111/j.1558-5646.1997.tb01457.x

Hansen, T. F. (2012). Adaptive landscapes and macroevolutionary dynamics. In The Adaptive Landscape in Evolutionary Biology (ed. E. Svensson and R. Calsbeek), pp. 205-226. London: Oxford University Press.

Hansen, T. F. (2014). Use and misuse of comparative methods in the study of adaptation. In Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology (ed. L. Z. Garamszegi), pp. 351-379. Berlin: Springer-Verlag.

Hansen, T. F. and Bartoszek, K. (2012). Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. Syst. Biol. 61, 413-425. doi:10.1093/sysbio/syr122

Hansen, T. F. and Martins, E. P. (1996). Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50, 1404-1417. doi:10.1111/j.1558-5646.1996.tb03914.x

Hansen, T. F. and Orzack, S. H. (2005). Assessing current adaptation and phylogenetic inertia as explanations of trait evolution: the need for controlled comparisons. Evolution 59, 2063-2072. doi:10.1554/05-088.1

Hansen, T. F., Pienaar, J. and Orzack, S. H. (2008). A comparative method for studying adaptation to a randomly evolving environment. Evolution 62, 1965-1977. doi:10.1111/j.1558-5646.2008.00412.x

Harmon, L. J., Losos, J. B., Davies, T. J., Gillespie, R. G., Gittleman, J. L., Jennings, W. B., Kozak, K. H., McPeek, M. A., Moreno-Roark, F., Near, T. J. et al. (2010). Early bursts of body size and shape evolution are rare in comparative data. Evolution 64, 2385-2396. doi:10.1111/j.1558-5646.2010.01025.x

Harvey, P. H. and Pagel, M. (1991). The Comparative Method in Evolutionary Biology. Oxford: Oxford University Press.

Herrel, A. and Bonneaud, C. (2012). Temperature dependence of locomotor performance in the tropical clawed frog, Xenopus tropicalis. J. Exp. Biol. 215, 2465-2470. doi:10.1242/jeb.069765

Hertz, P. E., Arima, Y., Harrison, A., Huey, R. B., Losos, J. B. and Glor, R. E. (2013). Asynchronous evolution of physiology and morphology in *Anolis* lizards. *Evolution* **67**, 2101-2113. doi:10.1111/evo.12072

Higham, T. E., Ferry, L. A., Schmitz, L., Irschick, D. J., Starko, S., Anderson, P. S. L., Bergmann, P. J., Jamniczky, H. A., Monteiro, L. R., Navon, D. et al. (2021). Linking ecomechanical models and functional traits to understand phenotypic diversity. *Trends Ecol. Evol.* **36**, 860-873. doi:10.1016/j.tree.2021.05.009

Hillebrand, H. (2004). On the generality of the latitudinal diversity gradient. *Am. Nat.* **163**, 192-211. doi:10.1086/381004

Ho, L. S. T. and Ané, C. (2013). Asymptotic theory with hierarchical autocorrelation: Ornstein–Uhlenbeck tree models. *Ann. Stat.* **41**, 957-981. doi:10.1214/13-AOS1105

Ho, L. S. T. and Ané, C. (2014a). Intrinsic inference difficulties for trait evolution with Ornstein–Uhlenbeck models. *Methods Ecol. Evol.* **5**, 1133-1146. doi:10.1111/2041-210X.12285

Ho, L. S. T. and Ané, C. (2014b). A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.* **63**, 397-408. doi:10.1093/sysbio/syu005

Huelsenbeck, J. P., Nielsen, R. and Bollback, J. P. (2003). Stochastic mapping of morphological characters. *Syst. Biol.* **52**, 131-158. doi:10.1080/10635150390192780

Huey, R. B. and Kingsolver, J. G. (1989). Evolution of thermal sensitivity of ectotherm performance. *Trends Ecol. Evol.* **4**, 131-135. doi:10.1016/0169-5347(89)90211-5

Huey, R. B. and Stevenson, R. D. (1979). Integrating thermal physiology and ecology of ectotherms: a discussion of approaches. *Am. Zool.* **19**, 357-366. doi:10.1093/icb/19.1.357

Hutchison, V. H., Whitford, W. G. and Kohl, M. (1968). Relation of body size and surface area to gas exchange in anurans. *Physiol. Zool.* **41**, 65-85. doi:10.1086/physzool.41.1.30158485

Ives, A. R., Midford, P. E. and Garland, T., Jr (2007). Within-species variation and measurement error in phylogenetic comparative methods. *Syst. Biol.* **56**, 252-270. doi:10.1080/10635150701313830

Janzen, D. H. (1967). Why mountain passes are higher in the tropics. *Am. Nat.* **101**, 233-249. doi:10.1086/282487

Jenkins, F. A. J. and Shubin, N. H. (1998). *Prosalirus bitis* and the anuran caudopelvic mechanism. *J. Vert. Paleontol.* **18**, 495-510. doi:10.1080/02724634.1998.10011077

Jetz, W. and Pyron, R. A. (2018). The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. *Nat. Ecol. Evol.* **2**, 850-858. doi:10.1038/s41559-018-0515-5

John-Alder, H. B., Morin, P. J. and Lawler, S. (1988). Thermal physiology, phenology, and distribution of tree frogs. *Am. Nat.* **132**, 506-520. doi:10.1086/284868

John-Alder, H. B., Barnhart, M. C. and Bennett, A. F. (1989). Thermal sensitivity of swimming performance and muscle contraction in northern and southern populations of tree frogs (*Hyla crucifer*). *J. Exp. Biol.* **142**, 357-372. doi:10.1242/jeb.142.1.357

Kellermann, V., Chown, S. L., Schou, M. F., Aitkenhead, I., Janion-Scheepers, C., Clemson, A., Scott, M. T. and Sgrò, C. M. (2019). Comparing thermal performance curves across traits: how consistent are they? *J. Exp. Biol.* **222**, jeb193433. doi:10.1242/jeb.193433

Kimura, M. T. (2004). Cold and heat tolerance of drosophilid flies with reference to their latitudinal distributions. *Oecologia* **140**, 442-449. doi:10.1007/s00442-004-1605-4

Knowles, T. W. and Weigl, P. D. (1990). Thermal dependence of anuran burst locomotor performance. *Copeia* **1990**, 796-802. doi:10.2307/1446445

Kozak, K. H. and Wiens, J. J. (2010). Niche conservatism drives elevational diversity patterns in Appalachian salamanders. *Am. Nat.* **176**, 40-54. doi:10.1086/653031

Krause, S., van Bodegom, P. M., Cornwell, W. K. and Bodelier, P. L. E. (2014). Weak phylogenetic signal in physiological traits of methane-oxidizing bacteria. *J. Evol. Biol.* **27**, 1240-1247. doi:10.1111/jeb.12401

Labra, A., Pienaar, J. and Hansen, T. F. (2009). Evolution of thermal physiology in *Liolaemus* lizards: adaptation, phylogenetic inertia, and niche tracking. *Am. Nat.* **174**, 204-220. doi:10.1086/600088

Lauder, G. V. (1982). Historical biology and the problem of design. *J. Theor. Biol.* **97**, 57-67. doi:10.1016/0022-5193(82)90276-4

Lauder, G. V. (1990). Functional morphology and systematics: studying functional patterns in an historical context. *Annu. Rev. Ecol. Syst.* **21**, 317-340. doi:10.1146/annurev.es.21.110190.001533

Lauder, G. V. (1991). Biomechanics and evolution: integrating physical and historical biology in the study of complex systems. In *Biomechanics in Evolution* (ed. J. M. V. Rayner and R. J. Wootton), pp. 1-19. Cambridge: Cambridge University Press.

Lauder, G. V. (1996). The argument from design. In *Adaptation* (ed. M. R. Rose and G. V. Lauder), pp. 55-91. San Diego, CA: Academic Press.

Lauder, G. V. (2003). The intellectual challenge of biomechanics and evolution. In *Vertebrate Biomechanics and Evolution* (eds. V. L. Bels, J.-P. Gasc and A. Casinos), pp. 319-325. Oxford: BIOS Scientific Publishers.

Lauder, G. V. and Liem, K. F. (1989). The role of historical factors in the evolution of complex organismal functions. In *Complex Organismal Functions: Integration and Evolution in Vertebrates* (ed. D. B. Wake and G. Roth), pp. 63-78. New York: John Wiley and sons.

Layne, J. R. J. and Romano, M. A. (1985). Critical thermal minima of *Hyla chrysoscelis, H. cinerea, H. gratiosa*, and natural hybrids (*H. cinerea* x *H. gratiosa*). *Herpetologica* **41**, 216-221.

Leroi, A. M., Rose, M. R. and Lauder, G. V. (1994). What does the comparative method reveal about adaptation? *Am. Nat.* **143**, 381-402. doi:10.1086/285609

Lewontin, R. C. (1978). Adaptation. *Sci. Am.* **239**, 212-230. doi:10.1038/scientificamerican0978-212

Li, T., Cao, P., Bei, Y.-J. and Du, W.-G. (2018). Latitudinal and temperature-dependent variation in embryonic development rate and offspring performance in a freshwater turtle. *Physiol. Biochem. Zool.* **91**, 673-681. doi:10.1086/694856

Logan, M. L., Cox, R. M. and Calsbeek, R. (2014). Natural selection on thermal performance in a novel thermal environment. *Proc. Natl. Acad. Sci. USA* **111**, 14165-14169. doi:10.1073/pnas.1404885111

Losos, J. B. (1999). Uncertainty in the reconstruction of ancestral character states and limitations on the use of phylogenetic comparative methods. *Anim. Behav.* **58**, 1319-1324. doi:10.1006/anbe.1999.1261

Ludwig, G., Sinsch, U. and Pelster, B. (2015). Behavioural adaptations of *Rana temporaria* to cold climates. *J. Therm. Biol.* **49-50**, 82-90. doi:10.1016/j.jtherbio.2015.02.006

Lutterschmidt, W. I. and Hutchison, V. H. (1997). The critical thermal maximum: history and critique. *Can. J. Zool.* **75**, 1561-1574. doi:10.1139/z97-783

Lynch, M. (1990). The rate of morphological evolution in mammals from the standpoint of the neutral expectation. *Evolution* **136**, 727-741. doi:10.1086/285128

MacLean, H. J., Sørensen, J. G., Kristensen, T. N., Loeschcke, V., Beedholm, K., Kellermann, V. and Overgaard, J. (2019). Evolution and plasticity of thermal performance: an analysis of variation in thermal tolerance and fitness in 22 *Drosophila* species. *Philos. Trans. R. Soc. Lond. B* **374**, 20180548. doi:10.1098/rstb.2018.0548

Maddison, W. P. and FitzJohn, R. G. (2015). The unsolved challenge to phylogenetic correlation tests for categorical characters. *Syst. Biol.* **64**, 127-136. doi:10.1093/sysbio/syu070

Mahoney, J. J. and Hutchison, V. H. (1969). Photoperiod acclimation and 24-h variations in the critical thermal maxima of a tropical and a temperate frog. *Oecologia* **2**, 143-161. doi:10.1007/BF00379157

Mannion, P. D., Upchurch, P., Benson, R. B. J. and Goswami, A. (2014). The latitudinal biodiversity gradient through deep time. *Trends Ecol. Evol.* **29**, 42-50. doi:10.1016/j.tree.2013.09.012

Martins, E. P. (1994). Estimating the rate of phenotypic evolution from comparative data. *Am. Nat.* **144**, 193-209. doi:10.1086/285670

Martins, E. P. (2000). Adaptation and the comparative method. *Trends Ecol. Evol.* **15**, 296-299. doi:10.1016/S0169-5347(00)01880-2

Martins, E. P. and Hansen, T. F. (1997). Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* **149**, 646-667. doi:10.1086/286013

Mendelson, J. R. I., Mulcahy, D. G., Williams, T. S. and Sites, J. W. J. (2011). A phylogeny and evolutionary natural history of Mesoamerican toads (Anura: Bufonidae: *Incilius*) based on morphology, life history, and molecular data. *Zootaxa* **3138**, 1-34. doi:10.11646/zootaxa.3138.1.1

Mendoza, E., Azizi, E. and Moen, D. S. (2020). What explains vast differences in jumping power within a clade? Diversity, ecology, and evolution of anuran jumping power. *Funct. Ecol.* **34**:1053-1063. doi:10.1111/1365-2435.13545

Mittelbach, G. G., Schemske, D. W., Cornell, H. V., Allen, A. P., Brown, J. M., Bush, M. B., Harrison, S. P., Hurlbert, A. H., Knowlton, N., Lessios, H. A. et al. (2007). Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecol. Lett.* **10**, 315-331. doi:10.1111/j.1461-0248.2007.01020.x

Moen, D. S. (2019). What determines the distinct morphology of species with a particular ecology? The roles of many-to-one mapping and trade-offs in the evolution of frog ecomorphology and performance. *Am. Nat.* **194**, E81-E95. doi:10.1086/704736

Moen, D. S. and Wiens, J. J. (2017). Microhabitat and climatic niche change explain patterns of diversification among frog families. *Am. Nat.* **190**, 29-44. doi:10.1086/692065

Moen, D. S., Smith, S. A. and Wiens, J. J. (2009). Community assembly through evolutionary diversification and dispersal in Middle American treefrogs. *Evolution* **63**, 3228-3247. doi:10.1111/j.1558-5646.2009.00810.x

Moen, D. S., Irschick, D. J. and Wiens, J. J. (2013). Evolutionary conservatism and convergence both lead to striking similarity in ecology, morphology and performance across continents in frogs. *Proc. R. Soc. B* **280**, 20132156. doi:10.1098/rspb.2013.2156

Moen, D. S., Morlon, H. and Wiens, J. J. (2016). Testing convergence versus history: convergence dominates phenotypic evolution for over 150 million years in frogs. *Syst. Biol.* **65**, 146-160. doi:10.1093/sysbio/syv073

Moen, D. S., Cabrera-Guzmán, E., Caviedes-Solis, I. W., González-Bernal, E. and Hanna, A. R. (2021a). Supplementary datasets, data analysis code, and R tutorials for: Phylogenetic analysis of adaptation in comparative physiology and biomechanics: overview and a case study of thermal physiology in treefrogs. Dryad Dataset. doi:10.5061/dryad.t4b8gtj2m.

Moen, D. S., Ravelojaona, R. N., Hutter, C. R. and Wiens, J. J. (2021b). Testing for adaptive radiation: a new approach applied to Madagascar frogs. Evolution (in press). doi:10.1111/evo.14328

Muñoz, M. M., Stimola, M. A., Algar, A. C., Conover, A., Rodriguez, A. J., Landestoy, M. A., Bakken, G. S. and Losos, J. B. (2014). Evolutionary stasis and lability in thermal physiology in a group of tropical lizards. Proc. R. Soc. B 281, 20132433. doi:10.1098/rspb.2013.2433

Nauwelaerts, S., Ramsay, J. and Aerts, P. (2007). Morphological correlates of aquatic and terrestrial locomotion in a semi-aquatic frog, Rana esculenta: no evidence for a design conflict. J. Anat. 210, 304-317. doi:10.1111/j.1469-7580. 2007.00691.x

Navas, C. A. (1996). Metabolic physiology, locomotor performance, and thermal niche breadth in Neotropical anurans. Physiol. Zool. 69, 1481-1501. doi:10.1086/ physzool.69.6.30164271

Navas, C. A., Gomes, F. R. and Carvalho, J. E. (2008). Thermal relationships and exercise physiology in anuran amphibians: integration and evolutionary implications. Comp. Biochem. Physiol. A Mol. Integr. Physiol. 151, 344-362. doi:10.1016/j.cbpa.2007.07.003

Near, J. C., Easton, D. P., Rutledge, P. S., Dickinson, D. P. and Spotila, J. R. (1990). Heat shock protein 70 gene expression in intact salamanders (Eurycea bislineata) in response to calibrated heat shocks and to high temperatures encountered in the field. J. Exp. Zool. 256, 303-314. doi:10.1002/jez.1402560310

Olson, M. E. (2021). The comparative method is not macroevolution: across-species evidence for within-species process. Syst. Biol. 70, 1272-1281. doi:10.1093/sysbio/syaa086

O'Meara, B. C. and Beaulieu, J. M. (2014). Modelling stabilizing selection: the attraction of Ornstein–Uhlenbeck models. In Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology (ed. L. Z. Garamszegi), pp. 381-393. Berlin: Springer-Verlag.

O'Meara, B. C., Ané, C., Sanderson, M. J. and Wainwright, P. C. (2006). Testing for different rates of continuous trait evolution using likelihood. Evolution 60, 922-933. doi:10.1111/j.0014-3820.2006.tb01171.x

Padilla, P., Ducret, V., Bonneaud, C., Courant, J. and Herrel, A. (2019). Acclimation temperature effects on locomotor traits in adult aquatic anurans (X. tropicalis and X. laevis) from different latitudes: possible implications for climate change. Conserv. Physiol. 7, coz019. doi:10.1093/conphys/coz019

Pennell, M. W. (2015). Book review: modern phylogenetic comparative methods and their application in evolutionary biology. Syst. Biol. 64, 161-163. doi:10.1093/ sysbio/syu075

Polato, N. R., Gill, B. A., Shah, A. A., Gray, M. M., Casner, K. L., Barthelet, A., Messer, P. W., Simmons, M. P., Guayasamin, J. M., Encalada, A. C. et al. (2018). Narrow thermal tolerance and low dispersal drive higher speciation in tropical mountains. Proc. Natl. Acad. Sci. USA 115, 12471-12476. doi:10.1073/ pnas.1809326115

Pontarp, M., Bunnefeld, L., Cabral, J. S., Etienne, R. S., Fritz, S. A., Gillespie, R., Graham, C. H., Hagen, O., Hartig, F., Huang, S. et al. (2019). The latitudinal diversity gradient: novel understanding through mechanistic eco-evolutionary models. Trends Ecol. Evol. 34, 211-223. doi:10.1016/j.tree.2018.11.009

Portik, D. M. and Papenfuss, T. J. (2015). Historical biogeography resolves the origins of endemic Arabian toad lineages (Anura: Bufonidae): evidence for ancient vicariance and dispersal events with the Horn of Africa and South Asia. BMC Evol. Biol. 15, 152. doi:10.1186/s12862-015-0417-y

Price, S. A., Friedman, S. T. and Wainwright, P. C. (2015). How predation shaped fish: the impact of fin spines on body form evolution across teleosts. Proc. R. Soc. B 282, 20151428. doi:10.1098/rspb.2015.1428

Putnam, R. W. and Bennett, A. F. (1981). Thermal dependence of behavioural performance of anuran amphibians. Anim. Behav. 29, 502-509. doi:10.1016/ S0003-3472(81)80111-X

Pyron, R. A. (2014). Biogeographic analysis reveals ancient continental vicariance and recent oceanic dispersal in amphibians. Syst. Biol. 63, 779-797. doi:10.1093/ sysbio/syu042

Ramsay, J. O., Hooker, G. and Graves, S. (2009). Functional Data Analysis with R and MATLAB. New York: Springer.

Reeve, H. K. and Sherman, P. W. (1993). Adaptation and the goals of evolutionary research. Q. Rev. Biol. 68, 1-32. doi:10.1086/417909

Renaud, J. M. and Stevens, E. D. (1983). The extent of long-term temperature compensation for jumping distance in the frog, Rana pipiens, and the toad, Bufo americanus. Can. J. Zool. 61, 1284-1287. doi:10.1139/z83-172

Revell, L. J. (2010). Phylogenetic signal and linear regression on species data. Methods Ecol. Evol. 1, 319-329. doi:10.1111/j.2041-210X.2010.00044.x

Rezende, E. L. and Bozinovic, F. (2019). Thermal performance across levels of biological organization. Philos. Trans. R. Soc. Lond. B 374, 20180549. doi:10.1098/rstb.2018.0549

Ricklefs, R. E. and Schluter, D. (1994). Species Diversity in Ecological Communities. Chicago: University of Chicago Press.

Roff, D. A. and Fairbairn, D. J. (2007). The evolution of trade-offs: where are we? J. Evol. Biol. 20, 433-447. doi:10.1111/j.1420-9101.2006.01255.x

Salisbury, B. A. and Kim, J. (2001). Ancestral state estimation and taxon sampling density. Syst. Biol. 50, 557-564. doi:10.1080/106351501750435103

Scales, J. A., King, A. A. and Butler, M. A. (2009). Running for your life or running for your dinner: what drives fiber-type evolution in lizard locomotor muscles? Am. Nat. 173, 543-553. doi:10.1086/597613

Schmid, W. D. (1982). Survival of frogs in low temperature. Science 215, 697-698. doi:10.1126/science.7058335

Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9, 671-675. doi:10.1038/ nmeth.2089

Seebacher, F., White, C. R. and Franklin, C. E. (2015). Physiological plasticity increases resilience of ectothermic animals to climate change. Nat. Clim. Change 5, 61-66. doi:10.1038/nclimate2457

Silvestro, D., Kostikova, A., Litsios, G., Pearman, P. B., Salamin, N. and Münkemüller, T. (2015). Measurement errors should always be incorporated in phylogenetic comparative analysis. Methods Ecol. Evol. 6, 340-346. doi:10.1111/ 2041-210X.12337

Smith, S. A., Stephens, P. R. and Wiens, J. J. (2005). Replicate patterns of species richness, historical biogeography, and phylogeny in holarctic treefrogs. Evolution 59, 2433-2450. doi:10.1111/j.0014-3820.2005.tb00953.x

Smith, S. A., Nieto-Montes de Oca, A., Reeder, T. W. and Wiens, J. J. (2007). A phylogenetic perspective on elevational species richness patterns in Middle American treefrogs: why so few species in lowland tropical rainforests? Evolution 61, 1188-1207. doi:10.1111/j.1558-5646.2007.00085.x

Snyder, G. K. and Weathers, W. W. (1975). Temperature adaptations in amphibians. Am. Nat. 109, 93-101. doi:10.1086/282976

Spicer, J. I., Morley, S. A. and Bozinovic, F. (2019). Physiological diversity, biodiversity patterns and global climate change: testing key hypotheses involving temperature and oxygen. Philos. Trans. R. Soc. Lond. B 374, 20190032. doi:10.1098/rstb.2019.0032

Storey, K. B. and Storey, J. M. (1992). Natural freeze tolerance in ectothermic vertebrates. Annu. Rev. Physiol. 54, 619-637. doi:10.1146/annurev.ph.54. 030192.003155

Streicher, J. W., Cox, C. L., Campbell, J. A., Smith, E. N. and de Sá, R. O. (2012). Rapid range expansion in the Great Plains narrow-mouthed toad (Gastrophryne olivacea) and a revised taxonomy for North American microhylids. Mol. Phylogenet. Evol. 64, 645-653. doi:10.1016/j.ympev.2012. 05.020

Sunday, J. M., Bates, A. E. and Dulvy, N. K. (2011). Global analysis of thermal tolerance and latitude in ectotherms. Proc. R. Soc. B 278, 1823-1830. doi:10.1098/rspb.2010.1295

Sunday, J., Bennett, J. M., Calosi, P., Clusella-Trullas, S., Gravel, S., Hargreaves, A. L., Leiva, F. P., Verberk, W. C. E. P., Olalla-Tárraga, M. A. and Morales-Castilla, I. (2019). Thermal tolerance patterns across latitude and elevation. Philos. Trans. R. Soc. Lond. B 374, 20190036. doi:10.1098/rstb.2019. 0036

Taylor, G. K. and Thomas, A. L. R. (2014). Evolutionary Biomechanics. Oxford: Oxford University Press.

Toljagić, O., Voje, K. L., Matschiner, M., Liow, L. H. and Hansen, T. F. (2018). Millions of years behind: slow adaptation of ruminants to grasslands. Syst. Biol. 67, 145-157. doi:10.1093/sysbio/syx059

Uyeda, J. C., Hansen, T. F., Arnold, S. J. and Pienaar, J. (2011). The million-year wait for macroevolutionary bursts. Proc. Natl. Acad. Sci. USA 108, 15908-15913. doi:10.1073/pnas.1014503108

Uyeda, J. C., Zenil-Ferguson, R. and Pennell, M. W. (2018). Rethinking phylogenetic comparative methods. Syst. Biol. 67, 1091-1109. doi:10.1093/ sysbio/syy031

van Berkum, F. H. (1988). Latitudinal patterns of the thermal sensitivity of sprint speed in lizards. Am. Nat. 132, 327-343. doi:10.1086/284856

Vogel, S. (2007). The emergence of comparative biomechanics. Integr. Comp. Biol. 47, 13-15. doi:10.1093/icb/icm004

von May, R., Catenazzi, A., Corl, A., Santa-Cruz, R., Carnaval, A. C. and Moritz, C. (2017). Divergence of thermal physiological traits in terrestrial breeding frogs along a tropical elevational gradient. Ecol. Evol. 7, 3257-3267. doi:10.1002/ ece3.2929

von May, R., Catenazzi, A., Santa-Cruz, R., Gutierrez, A. S., Moritz, C. and Rabosky, D. L. (2019). Thermal physiological traits in tropical lowland amphibians: vulnerability to climate warming and cooling. PLoS ONE 14, e0219759. doi:10.1371/journal.pone.0219759

Waldrop, L. D. and Rader, J. A. (2020). Melding modeling and morphology: a call for collaboration to address difficult questions about the evolution of form and function. Integr. Comp. Biol. 60, 1188-1192. doi:10.1093/icb/icaa132

Walker, J. A. (1998). Estimating velocities and accelerations of animal locomotion: a simulation experiment comparing numerical differentiation algorithms. J. Exp. Biol. 201, 981-995. doi:10.1242/jeb.201.7.981

Walton, B. M. (1993). Physiology and phylogeny: the evolution of locomotor energetics in hylid frogs. Am. Nat. 141, 26-50. doi:10.1086/285459

Whitehead, P. J., Puckridge, J. T., Leigh, C. M. and Seymour, R. S. (1989). Effect of temperature on jump performance of the frog *Limnodynastes tasmaniensis*. *Physiol. Zool.* **62**, 937-949. doi:10.1086/physzool.62.4.30157938

Whitford, W. G. (1973). The effects of temperature on respiration in the Amphibia. *Am. Zool.* **13**, 505-512. doi:10.1093/icb/13.2.505

Wiens, J. J. and Donoghue, M. J. (2004). Historical biogeography, ecology and species richness. *Trends Ecol. Evol.* **19**, 639-644. doi:10.1016/j.tree.2004.09.011

Wiens, J. J., Fetzner, J. W. J., Parkinson, C. L. and Reeder, T. W. (2005). Hylid frog phylogeny and sampling stratagies for speciose clades. *Syst. Biol.* **54**, 778-807. doi:10.1080/10635150500234625

Wiens, J. J., Graham, C. H., Moen, D. S., Smith, S. A. and Reeder, T. W. (2006). Evolutionary and ecological causes of the latitudinal diversity gradient in hylid frogs: treefrog trees unearth the roots of high tropical diversity. *Am. Nat.* **168**, 579-596. doi:10.1086/507882

Willig, M. R., Kaufman, D. M. and Stevens, R. D. (2003). Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Annu. Rev. Ecol. Evol. Syst.* **34**, 273-309. doi:10.1146/annurev.ecolsys.34.012103.144032

Wilson, R. S. (2001). Geographic variation in thermal sensitivity of jumping performance in the frog *Limnodynastes peronii*. *J. Exp. Biol.* **204**, 4227-4236. doi:10.1242/jeb.204.24.4227

Wilson, R. S. and Franklin, C. E. (2000). Absence of thermal acclimatory capacity of locomotor performance in adults of the frog *Limnodynastes peronii*. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **127**, 21-28. doi:10.1016/S1095-6433(00)00238-5

Wilson, R. S., James, R. S. and Johnston, I. A. (2000). Thermal acclimation of locomotor performance in tadpoles and adults of the aquatic frog *Xenopus laevis*. *J. Comp. Physiol. B* **170**, 117-124. doi:10.1007/s003600050266

Zug, G. R. (1985). Anuran locomotion: fatigue and jumping performance. *Herpetologica* **41**, 188-194.