Maternal behavioural thermoregulation facilitated evolutionary transitions from egg laying to live birth

Amanda K. Pettersen[1*], Nathalie Feiner[1], Daniel W.A. Noble[2], Geoffrey M. While[3], Tobias Uller[1#] & Charlie K. Cornwallis[1#]

1. Department of Biology, Lund University, Lund, 22 362, Sweden
2. Division of Ecology and Evolution, Research School of Biology, The Australian National University, Canberra, 2600, Australia
3. School of Natural Sciences, University of Tasmania, Hobart, 7005, Australia

*Corresponding author: amanda.pettersen@sydney.edu.au

#Joint senior author

Keywords: Viviparity, Oviparity, Thermal plasticity, Behavioural plasticity, Reproductive mode, Squamates, Lizards, Snakes

Abstract

Live birth is a key innovation that has evolved from egg laying ancestors over 100 times in reptiles. However, egg-laying lizards and snakes often have preferred body temperatures that are lethal to developing embryos and should prevent egg retention: how has viviparity repeatedly evolved in the face of this pervasive mismatch? We resolve this paradox by conducting phylogenetic analyses of adult and embryo thermal preferences across 224 species. Thermal mismatches between mothers and offspring are widespread but resolved by gravid females down-regulating their body temperature towards the thermal optimum of embryos. Importantly, this thermoregulatory behaviour evolved in ancestral egg-laying species before the evolution of live birth. Maternal thermoregulatory behaviour therefore bypasses the constraints imposed by a slowly evolving thermal physiology and is likely to have been a key requirement for repeated transitions to live birth.

Introduction

The evolution of live bearing is an important life-history transition in vertebrates (Doody and Moore, 2010; Lambert and Wiens, 2013; Shine, 1985). The ecological conditions that favour the transition from egg laying (oviparity) to live birth (viviparity) are relatively well understood, especially in reptiles, with particularly strong support for the adaptive value of viviparity in cool climates (Lambert and Wiens, 2013; Shine, 1995). By retaining embryos throughout development, mothers can buffer offspring from suboptimal nest temperatures, thereby ensuring faster development, higher hatching success, and increased offspring viability (Beuchat, 1988a; Le Henanff et al., 2013; Shine, 1995; Warner and Shine, 2007).

Despite the adaptive advantages of viviparity, its initial evolution from oviparous ancestors would seem unlikely. Embryos and adults of oviparous species are adapted to different thermal environments, with adult lizards and snakes typically having preferred body temperatures that exceed the upper lethal limit of embryos. For example, the average nest temperature of the Iberian emerald lizard, *Lacerta schreiberi*, is 24 °C, rarely exceeding 30 °C, while the preferred body temperature of females is 33 °C (Monasterio et al., 2013). Since embryos are well adapted to the temperatures typically experienced in the nest, they generally have limited capacity to develop at temperatures outside this range (Du et al., 2019; Noble et al., 2018b; Sanger et al., 2018). Prolonged exposure to other temperatures, including those that are optimal for adults, can result in offspring abnormalities and even death (Andrews et al., 2000; Braña and Ji, 2000; Noble et al., 2018b; Van Damme et al., 1992). A mismatch between thermal optima of embryos and females should therefore prevent mothers from retaining embryos throughout development, inhibiting evolutionary transitions to live birth (Beuchat, 1988b). Notwithstanding this apparent constraint, live birth has evolved over 100 times in squamate reptiles (Blackburn, 2006, 2015a).

How can the repeated evolution of viviparity be reconciled with the widespread thermal mismatch between embryos and adults in oviparous species? There are two potential scenarios by which viviparity can evolve in the presence of maternal-embryo thermal mismatches. First, viviparity may only evolve from oviparous ancestors where there is no mismatch, that is, where adult body temperature (P_{bt}) and embryo thermal optima (T_{opt}) are well aligned. Second, females may behaviourally adjust their body temperature when pregnant to close the gap between adult and embryo thermal optima, even when substantial mismatches in thermal preferences exist. Such plasticity may come at a cost to female performance, but temporarily shifting body temperature while gravid to match the thermal optima of embryos could eliminate thermal barriers to the evolution of viviparity.

Explaining how mismatches between female and offspring thermal optima have been resolved in viviparous lineages is challenging. Adult and embryo thermal preferences of the oviparous ancestors of viviparous lineages need to be reconstructed using available data on extant species, but data on thermal preferences across life stages are difficult to obtain. Often only some measures (preferred body temperature of oviparous adults, optimal incubation temperature of embryos, plasticity in female thermoregulation) are available for a subset of species, leading to patchily distributed data (Figure 1). Recent advances in phylogenetic methods help to resolve such issues, making it possible to infer ancestral states across multiple traits where information is non-overlapping across species (e.g., values of some traits are only available for some species and vice versa) (Hadfield and Nakagawa, 2010; Nakagawa, 2015). The accuracy of such analyses are, however, reliant on the amount of variation in traits explained by phylogeny (phylogenetic signal) as this determines the accuracy with which missing values can be predicted (Molina-Venegas et al., 2018).

We collated data on the thermal preferences of adult females (163 species) and embryos (51 species) of squamate reptiles and combined this with data on their reproductive

mode (Pyron and Burbrink, 2014) (Tables S1 & S2). First, we examined the ability of phylogenetic methods to overcome the patchily distributed data on the thermal biology and reproductive mode of reptiles. These analyses revealed that phylogenetic models were able to predict ancestral states with a high degree of accuracy with substantial missing data (Table S3). Second, we estimated the ancestral states of preferred body temperature of adult females and embryos to quantify the thermal conflict between mothers and embryos (Table S4-S6). Third, we tested the prediction that transitions to viviparity have been restricted to oviparous lineages with a low degree of maternal-embryo mismatch (Table S7-S9). This was assessed by calculating if viviparity has originated more frequently from oviparous ancestors where female *Pbt* overlaps with the thermal optima of embryos. Finally, we examined if females adjust their body temperature when gravid to reduce the mismatch between adult and embryo thermal optima (52 species), and if so, whether this behaviour evolved prior to the emergence of viviparity, or afterwards in response to such mismatch (Tables S9-S12).

Results and Discussion

Ancestral state reconstructions with missing data

There was high phylogenetic signal for all traits (phylogenetic heritability (phylo H^2) >0.7), indicating the potential to infer ancestral values of traits given data limitations (Tables S4-S6 & S10). We explicitly tested if models could accurately predict the state of tips and ancestral nodes using limited data in two ways. First, we compared estimates of ancestral states for reproductive mode obtained from an analysis of all data on reproductive mode ($n_{species}$ = 7831), to estimates from an analysis of only species with thermal data on a pruned phylogeny ($n_{species}$ = 224). We found that 91% of ancestral nodes were predicted to have the same reproductive mode. Second, we compared two models, one where all data on reproductive mode were included ($n_{species}$ = 7831), and one where data were restricted to only species with

thermal data ($n_{species}$= 224), with the remaining tips set as missing values ($n_{species}$ = 7608). These analyses showed that the data on 224 species could be used to correctly predict the reproductive modes of the 7608 'missing' species with 70% accuracy (Figure 3. All ancestral reconstructions are presented in Table S3). These results show that phylogenetic methods can be used accurately predict extensive amounts of missing data, and more specifically, species for which there were thermal data provided a general representation of the broader squamate phylogeny.

Wide-spread mismatch in thermal optima of oviparous mothers and their embryos

Across oviparous lizards and snakes, many species displayed significant disparity between adult female preferred body temperature (P_{bt}) and the temperature that maximizes hatching success (T_{opt}) (Multi-response BPMM (MR-BPMM): posterior mode (PM) of difference = 3.43, 95% Credible Interval (CI): -2.26, 10.78, *pMCMC* = 0.08. Figure 1 & 2; Table S4). The female P_{bt} of oviparous species is on average 5 °C higher than the embryonic T_{opt} (raw data mean ± SD P_{bt} 103 species: 32.41 ± 4.15 ºC; T_{opt} 47 species: 27.15 ± 2.12 ºC). Incubation experiments have shown that such a difference can significantly reduce hatching success, illustrating that exposure to those thermal conditions throughout development via the retention of offspring may jeopardise embryo survival (reviewed in (Noble et al., 2018b)). P_{bt} and T_{opt} of oviparous species were both estimated to evolve slowly (Figure 1; MR-BPMM: P_{bt} phylo H^2: 0.92, CI: 0.82, 0.98. T_{opt} phylo H^2: 0.92, CI: 0.68, 0.98. Table S5) and female P_{bt} and offspring T_{opt} only showed weak evidence of coevolving (Figure 2; phylogenetic correlation (MR-BPMM): PM = 0.72, CI: -0.36, 0.97, *pMCMC* = 0.18. Table S5), suggesting that this mismatch may be difficult to resolve.

Transitions to live birth occur despite mismatches between females and embryos

We found evidence that preferred body temperature was lower in viviparous, compared with oviparous, lineages (MR-BPMM: PM = -0.30, CI: -0.49, 0.00, *pMCMC* = 0.03; Figure 3; Table S6). This is consistent with the hypothesis that viviparity is an adaptation to cool climates (Lambert and Wiens, 2013; Shine, 1985; Watson et al., 2014). However, reconstructions of ancestral preferred body temperatures of adults indicated that in the majority of lineages, lower preferred body temperatures evolved after, not before, transitions to live bearing. This was evident from the lack of difference in P_{bt} between ancestors of egg-laying and live-bearing lineages (MR-BPMM, PM = 0.05, CI: -4.47, 2.77, *pMCMC* = 0.49; Table S7).

Ancestral reconstructions of P_{bt} and T_{opt} of oviparous species showed that transitions from egg laying to live bearing were not more likely to occur when adult and embryo thermal optima were similar: 10% of ancestors of live bearing species had aligned embryo and adult thermal optima compared to 17% of the ancestors of egg laying species (χ^2 = 0.91, df=1, P= 0.34. Tables S8-S9). As a result, in 90% of egg-laying ancestors of live-bearing species, there were significant mismatches in thermal optimum with embryos having lower preferred body temperature compared to adults (Tables S8-S9). There was only one case of T_{opt} being significantly higher than P_{bt} but this was in a lineage with viviparous ancestors and descendants (Tables S8-S9). In summary, only 10% of the transitions to viviparity are explained by embryo and adult thermal optima being matched, and the lower P_{bt} observed for viviparous taxa appears to have evolved after transitions to live birth, not before.

Female behavioural plasticity resolves the constraints imposed by thermal physiology

Next, we tested the hypothesis that the thermal mismatch between females and embryos during the evolution of viviparity is resolved by females adjusting their body temperature when gravid to meet the thermal requirements of their embryos (Beuchat, 1986). We

collected data on the preferred body temperature of gravid ($P_{bt\text{-}g}$) and non-gravid ($P_{bt\text{-}ng}$) adult females from 52 species of lizards and snakes from published field and experimental studies (viviparous n = 32, oviparous n = 20. Table S1). Using these data, an effect size of female behavioural plasticity, Hedges' *g* was calculated (positive values = higher P_{bt} when gravid, negative values = lower P_{bt} when gravid).

Across the 52 species, there was evidence that females with a high preferred body temperatures behaviourally reduce their body temperature when gravid more than females with low preferred body temperatures (Figure 4A; Table S10). This relationship is expected to be restricted to viviparous lineages if female behavioural plasticity evolves in response to selection for embryo retention. However, there were comparable negative phylogenetic correlations between Hedges' *g* and P_{bt} for both oviparous and viviparous taxa, showing that behavioural adjustment of body temperature when gravid occurs to a similar extent in both egg-laying and live-bearing species (Egg laying: PM = -0.69, CI: -0.98, -0.03, *pMCMC* = 0.04. Live bearing: PM = -0.91, CI: -0.98, -0.29, *pMCMC* = 0.01. Table S11). Furthermore, ancestral reconstructions of Hedges' *g* ('the adjustment of body temperature when gravid') showed that in lineages with thermal mismatches between adults and embryos, females adjusted their body temperature when gravid to a much greater extent than where thermal optima were aligned, in both the ancestors of egg-laying and live-bearing species (MR-BPMM: Egg-laying ancestors Hedges' *g* PM = -1.07, CI = -1.93, -0.59, *pMCMC* < 0.001. Live bearing ancestors Hedges' *g* PM = -1.09, CI = -2.73, -0.21, *pMCMC* = 0.004. Table S12). As a result, the inferred Hedges' *g* did not differ between the ancestors of oviparous and viviparous species (Figure 4B), suggesting that behavioural plasticity was present prior to the emergence of live birth.

Phylogenetic reconstruction of the transitions in reproductive mode in squamates is a contentious topic (King and Lee, 2015; Pyron, 2015; Pyron and Burbrink, 2014). Our re-analysis of published data on reproductive mode recovers similar results to King and Lee 2015 (see methods for information on importance of rate shifts in viviparity), but is not intended to add to, or resolve, this controversy, which requires greater phylogenetic resolution and/or data on reproductive mode. For example, the current phylogeny is based on 17 genes and there was substantial variation in the number of estimated origins of viviparity across the sampled trees ($n_{species}$ in tree = 224, median number of origins = 23, range = 12 to 29; see Table S9 for variation in estimated ancestral states across trees). Nevertheless, our results appear to refute the hypothesis that transitions from oviparity to viviparity have been limited to lineages where the preferred body temperature of adult females fall within the thermal tolerance of embryos.

Instead, our results suggest that an ancient and evolutionarily conserved ability of egg-laying females to adjust their body temperature during pregnancy provides a general resolution to the conflict over adult and embryo thermal optima removing barriers to the evolution of live bearing. The down-regulation of female body temperature in egg-laying species while gravid may appear surprising considering that most of these species lay their eggs within the early stages of development (commonly around the time of limb bud formation) (Andrews and Mathies, 2000). However, early developmental stages, involving gastrulation, neurulation and organogenesis, are potentially even more sensitive to temperature stress than later stages, which are predominantly associated with growth (Beuchat, 1988b; Sanger et al., 2018). The temperature sensitivity of early-stage embryos therefore suggests that resolutions to mother-offspring thermal conflicts are required in both egg-laying and live-bearing species. If true, the key innovation of live birth may owe its

evolutionary origin to mechanisms of behavioural temperature regulation put in place long before live birth emerged.

We suggest that the behavioural flexibility of female lizards and snakes enabled the gradual evolution of live birth by helping traverse potential fitness valleys associated with embryo retention. In many live-bearing lineages, temperature mismatches have subsequently been eliminated through the evolution of a lower preferred body temperature in females. However, extant species demonstrate that behavioural flexibility is frequently maintained in viviparous lineages that have colonised cool climates. Such behavioural flexibility provides females with the means to upregulate their thermoregulation in order to maintain embryos at temperatures significantly warmer than the external environment, and thus contributes to the adaptive value of viviparity in extant species (Andrews, 2000; Le Henanff et al., 2013; Shine, 1985; Uller and Olsson, 2010; Warner and Shine, 2007). Female thermoregulatory behaviour therefore appears to be a key adaptation enabling the constraints imposed by thermal mismatches between mothers and embryos to be overcome. In turn, this has likely facilitated both the evolution of live birth, and the expansion of reptiles into a variety of environments.

Methods

Literature search and data collection

To investigate the relationship between maternal behavioural plasticity and embryo thermal sensitivity, we used reproductive mode data from Pyron and Burbrink (2014), and collated existing data to form three datasets: 1) female preferred body temperature (P_{bt}) ("P_{bt} dataset"), 2) the temperature at which hatching success was maximised (T_{opt}) ("T_{opt} dataset"), and 3) female body temperature when gravid ($P_{bt\text{-}g}$) and not gravid ($P_{bt\text{-}ng}$) ("Hedges' *g* dataset"). Complete data are presented in Table S1. Data were collected for each of the variables from published literature. We conducted a literature search using ISI *Web of Science* (v.5.30) with search terms specific to each dataset.

For ancestral state reconstructions (see details below), we obtained parity mode (oviparous versus viviparous) data for 8006 squamate species, reconstructed from phylogenetic analysis (Pyron and Burbrink (2014); Data File 2). We included data only for species represented in the recent phylogenetic tree used in our main analyses (Tonini et al., 2016), resulting in a total of 7831 species.

The P_{bt} dataset provided preferred body temperatures (P_{bt}) of adult females measured at one point in time. We used the search terms 'body temperature*', along with one of the following: 'squamat*', 'lizard*', 'snake*' which yielded a total of 1075 papers. We only used data from studies where P_{bt} for females was stated explicitly unless pooled male/female data stated no significant effect of sex. Data on females that were described to be gravid, or data collected during the reproductive season were excluded. This provided a final dataset of female preferred body temperature for 163 species that was independent of the gravid and non-gravid measures used to calculate Hedges' *g* (live bearing: n = 61 and egg laying: n = 103, note P_{bt} data for *Zootoca vivipara* was available for both reproductive modes and both were included in the analyses – see below).

The T_{opt} dataset quantified the temperature that maximises embryonic survival (measured as hatching success; T_{opt}) in 51 egg laying species (Figure S1). Hatching success data were taken from the Reptile Developmental Database (Noble et al., 2018a) and supplemented with additional studies using the same method described for the P_{bt} dataset (search terms: 'temperature* AND incubat* AND hatch* OR surv*', along with one of the following: 'squamat*', 'lizard*', 'snake*'), yielding a total of 671 papers. We only included studies where three or more constant temperature treatments were used under controlled laboratory conditions, resulting in 661 papers being rejected due to irrelevance or overlap with the Reptile Development Database. This resulted in a final *Topt* dataset, consisting of 51 species, obtained from 81 studies.

Estimating embryo thermal optima (Topt)

Given that T_{opt} had a strong phylogenetic signal (Phylogenetic Heritability (H^2) = 0.95, 95% CI: 0.74 – 0.99; Table S4) we fitted a Bayesian Phylogenetic Mixed Effects Model (BPMM) to hatching success data for all species jointly. Including phylogenetic information allowed for the T_{opt} of each species to be estimated with greater accuracy and precision given that the range and number of temperatures across species varied (range: 10-40 ºC, mean number of temperatures per species ± SD: 7.62 ± 4.29). Compared to non-phylogenetic models, T_{opt} estimates produced from phylogenetic models (BPMM) showed smaller sampling error and avoided convergence problems in estimating model parameters.

Importantly, this approach was not used to estimate T_{opt} data for species without data, only to better predict T_{opt} values for species for which there were data. The T_{opt} BPMM model was run for 1,100,000 iterations with a burn-in of 100,000 iterations and thinning rate of 500, leaving us with 2,000 samples from the posterior distribution. Autocorrelation was low (lag values < 0.1) and trace plots showed chains mixed well for all parameters. Our model

included temperature as a fixed effect (estimating both a linear and quadratic slopes) and random slopes of temperature (linear and quadratic slopes) fitted at the phylogenetic level. From our BPMM we estimated T_{opt}, and its corresponding sampling variance, using the posterior distribution of fixed effects and best linear unbiased predictors (BLUPs) for the random slopes (linear and quadratic) for each species as follows:

$$Topt = -\frac{\left(T_f + T_{sp}\right)}{\left(2\left(T_f^2 + T_{sp}^2\right)\right)}$$

Where T_f and T_f^2 are the posterior linear and quadratic fixed effect estimates for temperature and T_{sp} and T_{sp}^2 are the posterior BLUPs for a given species extracted from the phylogenetic random slopes. Calculating T_{opt} using the posterior distribution of fixed and random effects meant that sampling error for a given species could be propagated to subsequent analyses (see below).

Estimating Hedges' g

For the final "Hedges' *g*" dataset, we collected mean, standard deviation and sample sizes for female preferred body temperatures when gravid ($P_{bt\text{-}g}$) and non-gravid ($P_{bt\text{-}ng}$) (independent from the P_{bt} dataset) from the same study/population. Within the 'title', 'abstract' or 'keywords' we used search terms 'body temperature* AND gravid* OR reproduct*' along with one of the following: 'squamat*', 'lizard*', 'snake*' which yielded a total of 721 papers. An effect size of female adjustment of body temperature when gravid for each species was measured as the standardised mean differences (Hedges' *g*) in preferred body temperatures between non-gravid and gravid females ($P_{bt\text{-}g}$ - $P_{bt\text{-}ng}$), adjusting for small sample sizes (Borenstein et al., 2009).

We included both field and laboratory measures of female body temperature, comprising studies that directly compared $P_{bt\text{-}g}$ and $P_{bt\text{-}ng}$. Only studies that provided both

sample size and error around mean of preferred body temperature were included. Of the 721 papers 648 papers were excluded due to irrelevance or insufficient data. Combined, this yielded a total of 73 studies published up to July 2019 from which effect sizes were calculated for 52 species (live bearing: n = 32 and egg laying: n = 20). Studies used artificial temperature gradients in the laboratory (n = 37) or measured preferred basking temperature in the field (n = 36). Laboratory studies generally measured body temperature in the same female during gestation ($P_{bt\text{-}g}$) and either before or after gestation ($P_{bt\text{-}ng}$) as repeated measures. In contrast, field studies often measured body temperature on a population during the reproductive season, comparing body temperatures in gravid and non-gravid females at a single time point. We examined the mean Hedges' *g* between laboratory or field studies and found there were no significant differences (PM (Oviparous) = -0.70, CI = -1.61, 0.60; PM (Viviparous) = 0.53, CI = -0.25, 1.28. Table S13. See Verification analyses below).

The final full dataset (Table S1) contained a total of 224 squamate species. Due to the inherent paucity of data, particularly for hatching success and body temperature in gravid females, there was little overlap of species between datasets – we found 39 species had data available for 2 datasets, while we were only able to obtain data for 4 species across all 3 datasets. Two species in our data are both oviparous and viviparous (*Saiphos equalis* and *Z. vivpara).* For *S. equalis* all data were from an oviparous population, but for *Z. vivipara* there were data for both oviparous and viviparous populations and we treated these different reproductive modes as being distinct in our dataset (Recknagel et al., 2018).

General Statistical Methods

We used Hidden Markov Models (HMMs) to reconstruct ancestral states of viviparity, Phylogenetic Ridge Regression (PRR) to test for rate shifts in continuous traits across the phylogenetic tree, and Bayesian Phylogenetic Mixed Effects Models with single (BPMM)

and multiple response variables (MR-BPMM) to estimate phylogenetic correlations between traits and reconstruct ancestral states of continuous variables. All analyses were conducted in R version 4.0.1(R Core Team, 2020).

Reconstructing ancestral values of viviparity using Hidden Rate Models (HRM)

The phylogenetic distribution of the evolutionary origins of viviparity in squamates remains highly debated (Blackburn, 2015a; Pyron, 2015). Our aim here was not to try to resolve this controversy. However, past literature has highlighted that the rate of evolution of viviparity varies across squamates, and this has important effects on ancestral reconstructions (Beaulieu and O'Meara, 2014; King and Lee, 2015; Wright et al., 2015). Not accounting for such rate variation has resulted in the ancestor of squamates being predicted to be viviparous and many reversals of viviparity to oviparity, both of which are thought to be unlikely (King and Lee, 2015; Lee and Shine, 2021; Pyron and Burbrink, 2015).

We therefore used hidden rate Markov models (HRMs), implemented in the R package 'corHMM' that can estimate variation in the rate of evolution of binary characters across phylogenies (Beaulieu et al., 2021). To do this, a number of different rate categories from one state (e.g. oviparity) to another state (e.g. viviparity) are pre-defined and then estimated across the phylogeny. The most likely number of rate categories can be identified by comparing AIC values across models with different numbers of pre-defined rate categories.

We found on the trimmed phylogeny (224 species) that AIC values were lowest when there were 2 rate categories (See R script '4. PBT models.R section'). This indicated that in two clades, transitions to viviparity occurred at a higher rate than in other parts of the phylogeny (Figure S2). This model produced ancestral estimates that are consistent with the current predominant view of viviparity evolution across squamates (Blackburn, 2015b; King

and Lee, 2015): a root state of oviparity and relatively few reversals of viviparity to oviparity compared to the transitions from oviparity to viviparity (Table S3).

Estimates of ancestral states from HRMs were used to identify transitions between oviparity and viviparity by classifying nodes in the following way: 1) oviparous with only oviparous descendants (oviparous to oviparous); 2) viviparous with only viviparous descendants (viviparous to viviparous); 3) oviparous with at least one viviparous descendant (oviparous to viviparous); and 4) viviparous with at least one oviparous descendant (viviparous to oviparous).

Testing for rate shifts in T_{opt} and P_{bt} using Phylogenetic Ridge Regression (PRR)

We examined whether there was variation in the rates of evolution of T_{opt} and P_{bt} across the trimmed phylogeny (224 species) using phylogenetic ridge regression implemented in the R package 'RRphylo' (Castiglione et al., 2018). We found no evidence that there were rate shifts in T_{opt}, but for P_{bt}, two shifts were identified, one lower and one higher (Figure S2). We examined how important accounting for such rate variation was for reconstructing ancestral values of P_{bt} by comparing estimates from PRR models, both including and excluding clades with different rates, to estimates gained from BPMMs.

We found that ancestral estimates of P_{bt} from PRR models were correlated to estimates from BPMMs and this did not change according to whether clades with rates shifts were included or not (Pearson's correlation coefficient (*r*): included = 0.72, excluded = 0.72. See R script '4. PBT models.R section'). Given rate shifts had minimal impact on estimates of ancestral states, we used BPMMs because they: 1) more accurately predicted tip states than PRRs (BPMMs r = 0.84, PRRs r = 0.60); 2) allow missing data; 3) can incorporate sampling variances associated with response variables; 4) can be used to calculate phylogenetic correlations; and 5) can produce distributions of estimates (posterior samples)

for each node that enable significant thermal mismatches between embryos and adults to be estimated (see below).

Bayesian Phylogenetic Mixed Effects Models (BPMM)

We implemented BPMMs in *R* with the *MCMCglmm* package (Hadfield, 2010). T_{opt}, P_{bt}, Hedges' *g* were modelled with Gaussian error distributions. The probability of viviparity was modelled using a binary error distribution with a logit link function. For some species there were multiple estimates of Hedges' *g*, which was accounted for in two ways. In the main analyses (see specific analyses below for details) where multi-response models were used to examine relationships between traits, a single data point was included for each species. Therefore, a weighted mean for Hedges' *g* for each species was calculated. For verification analyses of Hedges' *g*, where other traits were not included, species was included as a random effect to account for multiple data points per species.

The random effect *animal*, with a covariance matrix derived from the phylogenetic tree was included in all models (Hadfield, 2010). We calculated the phylogenetic signal (equivalent to heritability, H^2, in the terminology of *MCMCglmm*) for each trait as the variance explained by *animal* relative to total variance (animal and residual). In the case of binary variables (reproductive mode) the residual variance is not identifiable and we therefore used the intraclass coefficient of variation (ICC: phylogenetic variance / (phylogenetic variance + residual variance + $pi^{2/3}$) * 100) to examine phylogenetic signature (Nakagawa et al., 2017).

Multi-response BPMMs (MR-BPMMs) can also be fitted with *MCMCglmm* that allow the phylogenetic and within-species (residual) correlations between traits to be estimated. Correlations between traits (e.g., A & B) were calculated as:

$$\frac{cov(A,B)}{\sqrt{(var(A) \cdot var(B))}}$$

Missing data across species

BPMMs permit missing data in response variables which was crucial given the patchy distribution of the data (Figure 1: T_{opt} = 51 species; P_{bt} = 163 species; Hedges' g = 52 species). The accuracy with which missing data is predicted is related to the phylogenetic signal in traits and the strength of phylogenetic correlations between traits (Molina-Venegas et al., 2018). All traits had high phylogenetic signal (phylogenetic heritability >70%) producing high correspondence between raw and predicted values (Figure S2; See also *"The ability of phylogenetic methods to predict missing data in reproductive mode"*). As a result, our BPMMs enabled us to deal with the fact that not all traits have been measured in all species.

Model convergence, prior settings and characterisation of posterior distributions

Non-informative uniform priors were used for fixed effects and inverse-Wishart priors for random effects (V = 1, nu = 0.002; (Hadfield, 2010)) apart from models with binary response variables (e.g. reproductive mode). For binary variables the residual variance was fixed to 1 and we specified a fixed effect prior that is relatively flat on the logit scale (mu = 0, V = residual variance x $\pi^{2/3}$) and parameter-expanded random effect priors (V = 1, nu = 0.002, alpha.mu = 0, alpha.V = 1000). To examine model convergence we ran three independent MCMC chains and examined autocorrelation, which was low (lag values < 0.1), trace plots, which showed chains mixed well, and Gelman and Rubin's convergence diagnostic that models converged (potential scale reduction factors were all below 1.1: R function gelman.diag (Brooks and Gelman, 1998). The convergence of models of the probability of viviparity (binary response variable) occurred more slowly and therefore were run for longer (burn-in = 1000000, iterations = 6000000, thinning = 5000). All other models used the run

settings specified in the data imputation section. Posterior distributions of all parameters were characterised using modes and 95% credible intervals (CIs) throughout. Effects were regarded as significant where CIs did not span 0. *pMCMC* (number of iterations above or below 0 / total number of iterations) are also presented to facilitate general interpretation.

Imputation for missing sampling variances

The accuracy of measures of T_{opt}, P_{bt}, and Hedges' g varied across species due to study design and sample sizes which can be accounted for by weighting data points by their inverse sampling variance using the 'mev' term in *MCMCglmm.* However, missing values in sampling variances are not permitted in *MCMCglmm.* As data on the error and sample size was missing for T_{opt}, P_{bt} and Hedges' g it would not have been possible to account for sampling error in our analyses without drastically reducing the size of our dataset. Consequently, we used multiple imputation with predictive mean matching in the *mice* package in R to impute missing error and sample sizes (Buuren and Groothuis-Oudshoorn, 2011). Samples sizes were not available for reproductive mode, but it is typically invariant within populations leading to minimal measurement error. Therefore, the mev term for reproductive mode was specified as 0.

To incorporate uncertainty in imputations, 20 complete datasets were generated, and all analyses were conducted by sampling across these datasets. Each model sampled through the 20 datasets 75 times (1500 sampling events) for 2000 iterations with only the last iteration being saved. Models that included the probability of viviparity were run for 5000 iterations, instead of 2000, to reduce autocorrelation amongst posterior samples that is common for binary variables. Estimates from the last iteration of each sampling event i were used as the starting parameter values for the next $i + 1$. This led to a posterior sample of 1500 iterations, the first 500 iterations were discarded as a burn-in and the remaining 1000 (50 per dataset)

were used to estimate parameters. Pooling of posterior distributions from model parameters from across each of the $n = 20$ datasets enabled imputation uncertainty in sampling variances to be accounted for.

Phylogeny and accounting for phylogenetic uncertainty

We used a recent phylogeny of 9754 squamates (Tonini et al., 2016). For the main analyses we used 1500 posterior trees to account for phylogenetic uncertainty pruned to the 224 species with thermal data. For BPMMs each model sampled through the 1500 trees using the same procedure as described for sampling across imputed datasets. Each of the 1500 posterior samples was therefore obtained using a different tree (for similar procedure see (Ross et al., 2013)). Pooling of posterior distributions from model parameters across trees allows for phylogenetic uncertainty to be taken into account (Pagel et al., 2004). For reconstructing ancestral states of viviparity, HRMs were repeated across the 1000 trees used for estimation in BPMMs. For figures and to examine the sensitivity of ancestral reconstructions to missing data, we used the maximum clade credibility tree provided by Tonini et al., (2016).

Phylogenetic uncertainty and predicting ancestral states

Quantifying discrepancies between embryo and adult thermal optima in relation to transitions in reproductive mode requires summarising posterior distributions of estimates of T_{opt} and P_{bt} for each node and relating it to its transition category (oviparous to oviparous, viviparous to viviparous, oviparous to viviparous, viviparous to viviparous). One complication is that for each node the predicted transition category can vary across trees due to differences in topology and internal tree structure. Discrepancies between T_{opt} and P_{bt} can be estimated for each transition category for each node, but this becomes problematic when some transition

categories for some nodes are rare as it results in few posterior samples to estimate discrepancies. To circumvent this problem, we estimated the ancestral states for each trait for each node in each tree (Table S9). Each node was then classified according to the most frequently predicted transition category and related to posterior distributions of T_{opt}, P_{bt} and Hedges' *g* summarised across all trees.

Specific Statistical Analyses

Ancestral state reconstructions with missing data

We examined how well models predicted ancestral values of reproductive mode with missing tip data using HRMs. We first compared the ancestral states of nodes predicted using all available data on reproductive mode ($n_{species}$ = 7831, Table S2) to the predicted states for same nodes that were obtained using only the trimmed tree and data on species in the thermal dataset ($n_{species}$ = 224, Table S1). Second, we examined the accuracy with which ancestral nodes could be predicted on the phylogeny of 7831 species using only reproductive mode data from the 224 species with thermal data. The predicted ancestral states from both these analyses can be found in Table S3.

Discrepancies between embryo and adult thermal optima and the evolution of live birth

We conducted six main analyses: **1)** Phylogenetic signature in P_{bt} and T_{opt} was estimated using a MR-BPMM of P_{bt} and T_{opt} with heterogenous phylogenetic and residual covariance matrices fitted as random effects (R code model M1.1. Table S2). **2)** The phylogenetic correlation between P_{bt} and T_{opt} was estimated using a MR-BPMM with unstructured phylogenetic and residual covariance matrices fitted as random effects (R code model M1.4. Table S3). **3)** The phylogenetic correlation between the probability of live bearing and P_{bt} was estimated using a MR-BPMM with unstructured phylogenetic and residual covariance

matrices fitted as random effects (R code model M2.1. Table S6). **4)** To quantify the number of ancestral egg-laying with and without mismatched adult female and embryo thermal optima a two-step approach was used. The ancestral states of reproductive mode were first estimated for each node using HRMs (R code model M3_corHMM). Next a MR-BPPM with P_{bt}, T_{opt} and Hedges' *g* as response variables was used (R code model M3.1) to reconstruct ancestral states for each trait. From this model mismatches in thermal optima (CI of $P_{bt} - T_{opt}$ not overlapping 0) and levels of female behavioural plasticity were estimated (Table S9 & S12). **5)** To examine if female plasticity differed between the ancestors of oviparous and viviparous lineages, with and without mismatched mother-offspring thermal optima, we examined estimates of Hedges' *g* in relation to predicted states of reproductive mode (R code model M3.1. Table S7 & S12). **6)** The phylogenetic correlation between P_{bt} and Hedges' *g* was estimated across all species (R code model M3.1. Table S10), as well as for egg-laying and live-bearing species separately using an MR-BPMM with separate unstructured phylogenetic and residual covariance matrices for each reproductive mode (R code model M4.1. Table S11).

Verification analyses

Checking for differences in Hedges' g between laboratory and field studies

Differences in the ability of laboratory versus field studies to estimate plasticity in female body temperature when gravid was examined using a BPMM of Hedges' *g* with study type as a fixed effect (R code model M5.4. Table S13).

Data and code availability

All data and code are publicly available at: https://doi.org/10.17605/OSF.IO/JT28V

Author Contributions: Conceptualization, N.F., T.U. and C.K.C; Methodology, D.W.A.N. and C.K.C.; Formal analysis, A.K.P., D.W.A.N. and C.K.C; Investigation, A.K.P., D.W.A.N., G.M.W. and T.U.; Data curation, A.K.P., D.W.A.N., T.U. and C.K.C; Writing – original draft, A.K.P, N.F., T.U. and C.K.C; Writing – Reviewing & Editing, A.K.P., N.F., D.W.A.N., G.M.W., T.U and C.K.C; Visualization, A.K.P., D.W.A.N. and C.K.C.; Supervision, T.U. and C.K.C..; Funding acquisition, A.K.P. and T.U.

Competing Interest Statement: The authors declare no competing interests.

References

Andrews RM. 2000. Evolution of viviparity in squamate reptiles (Sceloporus spp.): a variant of the cold-climate model. *Journal of Zoology* **250**:243–253. doi:10.1111/j.1469-7998.2000.tb01075.x

Andrews RM, Mathies T. 2000. Natural History of Reptilian Development: Constraints on the Evolution of Viviparity. *BioScience* **50**:227–238. doi:10.1641/0006-3568(2000)050[0227:NHORDC]2.3.CO;2

Andrews RM, Mathies T, Warner DA. 2000. Effect of Incubation Temperature on Morphology, Growth, and Survival of Juvenile Sceloporus undulatus. *Herpetological Monographs* **14**:420–431. doi:10.2307/1467055

Beaulieu J, O'Meara B, Oliver J, Boyko J. 2021. corHMM: Hidden Markov Models of Character Evolution.

Beaulieu JM, O'Meara BC. 2014. Hidden Markov Models for Studying the Evolution of Binary Morphological Characters In: Garamszegi LZ, editor. Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice. Berlin, Heidelberg: Springer. pp. 395–408. doi:10.1007/978-3-662-43550-2_16

Beuchat CA. 1988a. Temperature effects during gestation in a viviparous lizard. *Journal of Thermal Biology* **13**:135–142. doi:10.1016/0306-4565(88)90024-1

Beuchat CA. 1988b. Temperature effects during gestation in a viviparous lizard. *Journal of Thermal Biology* **13**:135–142. doi:10.1016/0306-4565(88)90024-1

Beuchat CA. 1986. Reproductive Influences on the Thermoregulatory Behavior of a Live-Bearing Lizard. *Copeia* **1986**:971–979. doi:10.2307/1445294

Blackburn DG. 2015a. Evolution of viviparity in squamate reptiles: Reversibility reconsidered. *J Exp Zool B Mol Dev Evol* **324**:473–486. doi:10.1002/jez.b.22625

Blackburn DG. 2015b. Evolution of viviparity in squamate reptiles: Reversibility reconsidered. *J Exp Zool B Mol Dev Evol* **324**:473–486. doi:10.1002/jez.b.22625

Blackburn DG. 2006. Squamate Reptiles as Model Organisms for the Evolution of Viviparity. *Herpetological Monographs* **20**:131–146.

Borenstein M, Hedges LV, Higgins JP, Rothstein HR. 2009. Effect Sizes Based on MeansIntroduction to Meta-Analysis. Chichester, UK: John Wiley & Sons, Ltd. pp. 21–32. doi:10.1002/9780470743386.ch4

Braña F, Ji X. 2000. Influence of incubation temperature on morphology, locomotor performance, and early growth of hatchling wall lizards (Podarcis muralis). *J Exp Zool* **286**:422–433. doi:10.1002/(sici)1097-010x(20000301)286:4<422::aid-jez10>3.0.co;2-d

Brooks SP, Gelman A. 1998. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics* **7**:434–455.

Buuren S van, Groothuis-Oudshoorn K. 2011. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* **45**:1–67. doi:10.18637/jss.v045.i03

Castiglione S, Tesone G, Piccolo M, Melchionna M, Mondanaro A, Serio C, Di Febbraro M, Raia P. 2018. A new method for testing evolutionary rate variation and shifts in phenotypic evolution. *Methods in Ecology and Evolution* **9**:974–983. doi:10.1111/2041-210X.12954

Doody JS, Moore JA. 2010. Conceptual model for thermal limits on the distribution of reptiles. *Herpetological Conservation and Biology* **5**:283–289.

Du W-G, Shine R, Ma L, Sun B-J. 2019. Adaptive responses of the embryos of birds and reptiles to spatial and temporal variations in nest temperatures. *Proceedings of the Royal Society B: Biological Sciences* **286**:20192078. doi:10.1098/rspb.2019.2078

Hadfield JD. 2010. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software* **33**:1–22. doi:10.18637/jss.v033.i02

Hadfield JD, Nakagawa S. 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology* **23**:494–508. doi:10.1111/j.1420-9101.2009.01915.x

King B, Lee MSY. 2015. Ancestral state reconstruction, rate heterogeneity, and the evolution of reptile viviparity. *Syst Biol* **64**:532–544. doi:10.1093/sysbio/syv005

Lambert SM, Wiens JJ. 2013. Evolution of viviparity: a phylogenetic test of the cold-climate hypothesis in Phrynosomatid lizards. *Evolution* **67**:2614–2630.

Le Henanff M, Meylan S, Lourdais O. 2013. The sooner the better: reproductive phenology drives ontogenetic trajectories in a temperate squamate (Podarcis muralis). *Biol J Linn Soc* **108**:384–395. doi:10.1111/j.1095-8312.2012.02005.x

Lee MSY, Shine R. 2021. Reptilian Viviparity and Dollo's Law 11.

Molina-Venegas R, Moreno-Saiz JC, Parga IC, Davies TJ, Peres-Neto PR, Rodríguez MÁ. 2018. Assessing among-lineage variability in phylogenetic imputation of functional trait datasets. *Ecography* **41**:1740–1749. doi:https://doi.org/10.1111/ecog.03480

Monasterio C, Shoo LP, Salvador A, Iraeta P, Díaz JA. 2013. High temperature constrains reproductive success in a temperate lizard: implications for distribution range limits and the impacts of climate change. *Journal of Zoology* **291**:136–145. doi:10.1111/jzo.12057

Nakagawa S. 2015. Missing data: mechanisms, methods, and messagesEcological Statistics. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199672547.003.0005

Nakagawa S, Johnson PCD, Schielzeth H. 2017. The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of The Royal Society Interface* **14**:20170213. doi:10.1098/rsif.2017.0213

Noble DWA, Stenhouse V, Riley JL, Warner DA, While GM, Du W-G, Uller T, Schwanz LE. 2018a. A comprehensive database of thermal developmental plasticity in reptiles. *Sci Data* **5**:1–7. doi:10.1038/sdata.2018.138

Noble DWA, Stenhouse V, Schwanz LE. 2018b. Developmental temperatures and phenotypic plasticity in reptiles: a systematic review and meta-analysis. *Biological Reviews* **93**:72–97. doi:10.1111/brv.12333

Pagel M, Meade A, Barker D. 2004. Bayesian Estimation of Ancestral Character States on Phylogenies. *Systematic Biology* **53**:673–684. doi:10.1080/10635150490522232

Pyron RA. 2015. Advancing perspectives on parity-mode evolution. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **324**:562–563. doi:10.1002/jez.b.22644

Pyron RA, Burbrink FT. 2015. Contrasting models of parity-mode evolution in squamate reptiles. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **324**:467–472. doi:10.1002/jez.b.22593

Pyron RA, Burbrink FT. 2014. Early origin of viviparity and multiple reversions to oviparity in squamate reptiles. *Ecology Letters* **17**:13–21. doi:https://doi.org/10.1111/ele.12168

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Recknagel H, Kamenos NA, Elmer KR. 2018. Common lizards break Dollo's law of irreversibility: Genome-wide phylogenomics support a single origin of viviparity and

re-evolution of oviparity. *Molecular Phylogenetics and Evolution* **127**:579–588. doi:10.1016/j.ympev.2018.05.029

Ross L, Gardner A, Hardy N, West SA. 2013. Ecology, Not the Genetics of Sex Determination, Determines Who Helps in Eusocial Populations. *Current Biology* **23**:2383–2387. doi:10.1016/j.cub.2013.10.013

Sanger TJ, Kyrkos J, Lachance DJ, Czesny B, Stroud JT. 2018. The effects of thermal stress on the early development of the lizard Anolis sagrei. *Journal of Experimental Zoology Part A: Ecological and Integrative Physiology* **329**:244–251. doi:10.1002/jez.2185

Shine R. 1995. A new hypothesis for the evolution of viviparity in reptiles. *The American Naturalist* **145**:809–823.

Shine R. 1985. The evolution of viviparity in reptiles: An ecological analysisBiology of the Reptilia, Development B. New York: John Wiley & Sons, Ltd. pp. 605–694.

Tonini JFR, Beard KH, Ferreira RB, Jetz W, Pyron RA. 2016. Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. *Biological Conservation*, Advancing reptile conservation: Addressing knowledge gaps and mitigating key drivers of extinction risk **204**:23–31. doi:10.1016/j.biocon.2016.03.039

Uller T, Olsson M. 2010. Offspring size and timing of hatching determine survival and reproductive output in a lizard. *Oecologia* **162**:663–671. doi:10.1007/s00442-009-1503-x

Van Damme R, Bauwens D, Braña F, Verheyen RF. 1992. Incubation Temperature Differentially Affects Hatching Time, Egg Survival, and Hatchling Performance in the Lizard Podarcis muralis. *Herpetologica* **48**:220–228.

Warner DA, Shine R. 2007. Fitness of juvenile lizards depends on seasonal timing of hatching, not offspring body size. *Oecologia* **154**:65–73. doi:10.1007/s00442-007-0809-9

Watson CM, Makowsky R, Bagley JC. 2014. Reproductive mode evolution in lizards revisited: updated analyses examining geographic, climatic and phylogenetic effects support the cold-climate hypothesis. *Journal of Evolutionary Biology* **27**:2767–2780. doi:10.1111/jeb.12536

Wright AM, Lyons KM, Brandley MC, Hillis DM. 2015. Which came first: The lizard or the egg? Robustness in phylogenetic reconstruction of ancestral states. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **324**:504–516. doi:10.1002/jez.b.22642

Figures

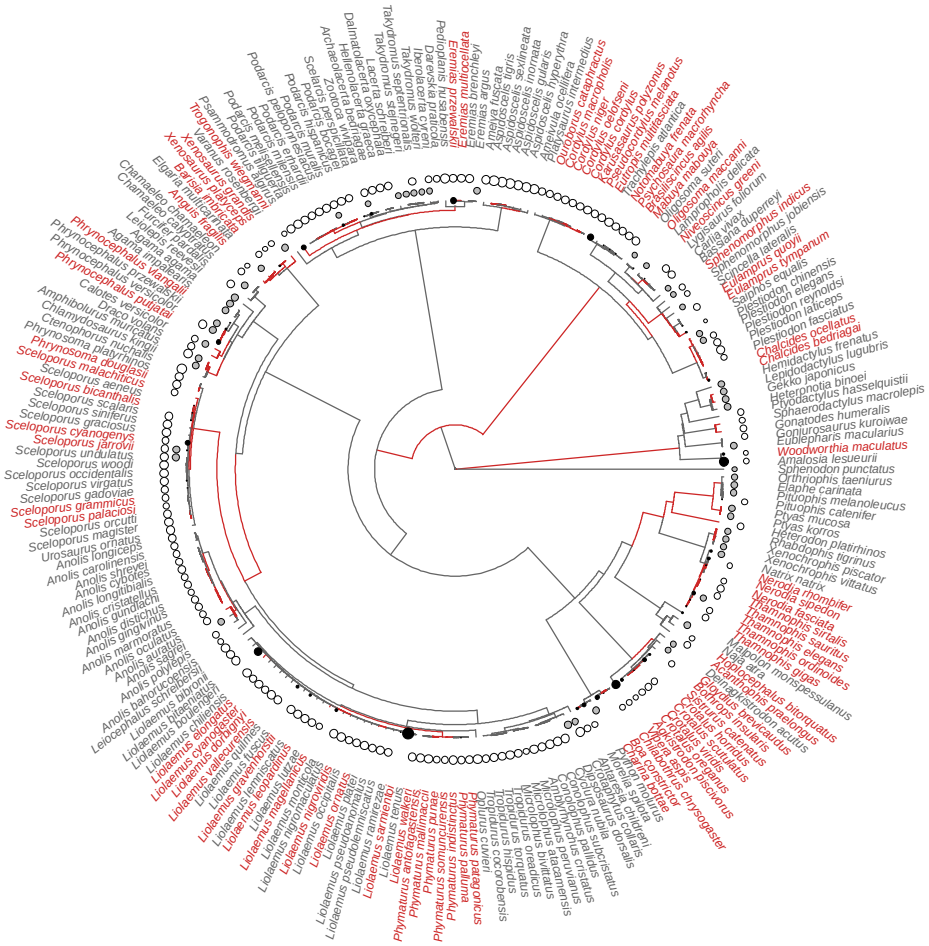

Figure 1. Variation in reproductive mode, preferred body temperature and plasticity in female thermo-regulatory behaviour across 224 species of squamate reptile (lizards and snakes). Tips and branches are coloured according to reproductive modes (red = live bearing/viviparous, grey = egg laying/oviparous; branch colours represent predicted ancestral values; Table S9). The size of white circles corresponds to the preferred body temperature of female squamates (P_{bt} dataset, 163 species), the size of grey circles to the optimal temperature for embryo survival in oviparous species (T_{opt} dataset, 51 species), and the size of the black circles to the discrepancy between preferred body temperature in gravid versus non-gravid females (Hedges' *g* dataset; absolute values, 52 species).

647

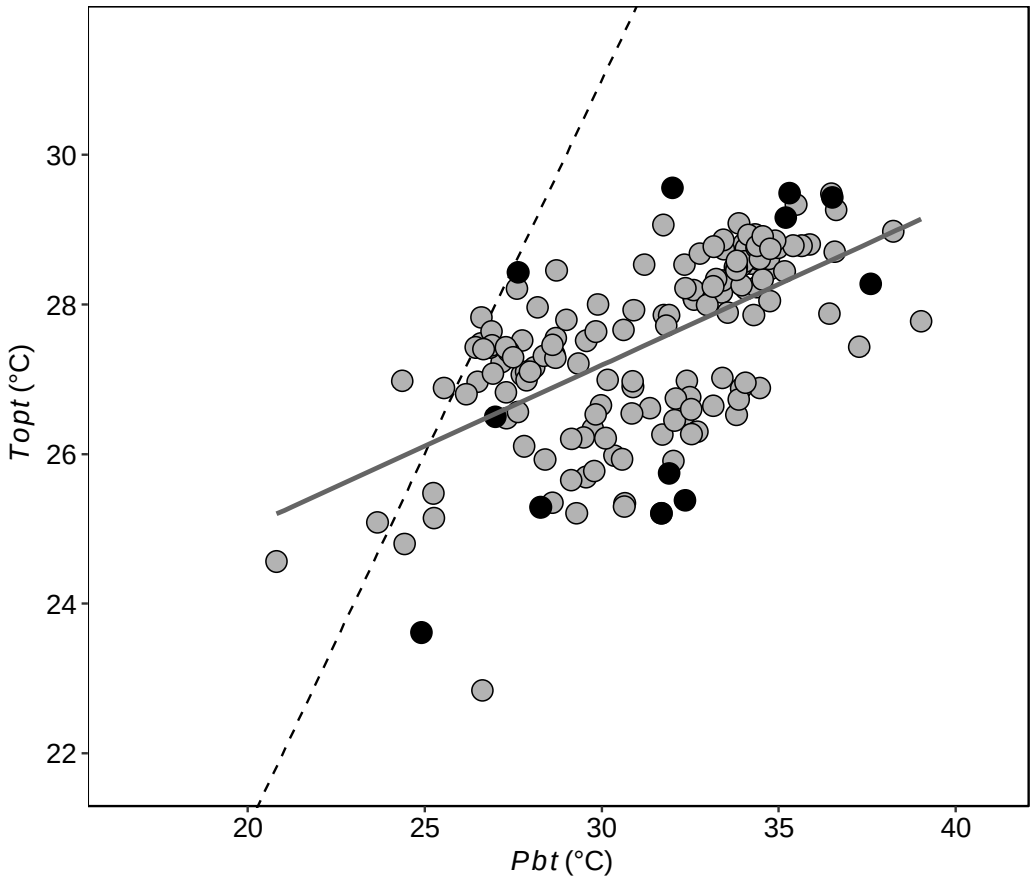

Figure 2. Mismatches in the thermal optima of mothers and embryos across 147 oviparous species of lizards and snakes. The relationship between female preferred body temperature (P_{bt}) and the temperature that maximizes hatching success of embryos (T_{opt}) predicted by a multi-response Bayesian Phylogenetic Mixed Model across egg laying species (grey data points, regression lines ± 95% confidence intervals are plotted). Raw data for 12 species with both P_{bt} and T_{opt} data shown by black data points. Despite high phylogenetic signature in P_{bt} and T_{opt}, there was only weak evidence for co-evolution between the two traits (Table S5). Deviations from the dotted line (1:1 relationship) indicate the magnitude of the mismatch between adult and embryo thermal optima. Corresponding measurements of T_{opt} and P_{bt} were not available for all oviparous species (total $n_{oviparous}$ = 147, n_{Pbt} = 103; n_{Topt} = 47; n_{Pbt} & n_{Topt} = 10), hence model predictions rather than raw data are shown (for plot of raw data as well as correspondence between predicted and raw data values see Figure S3).

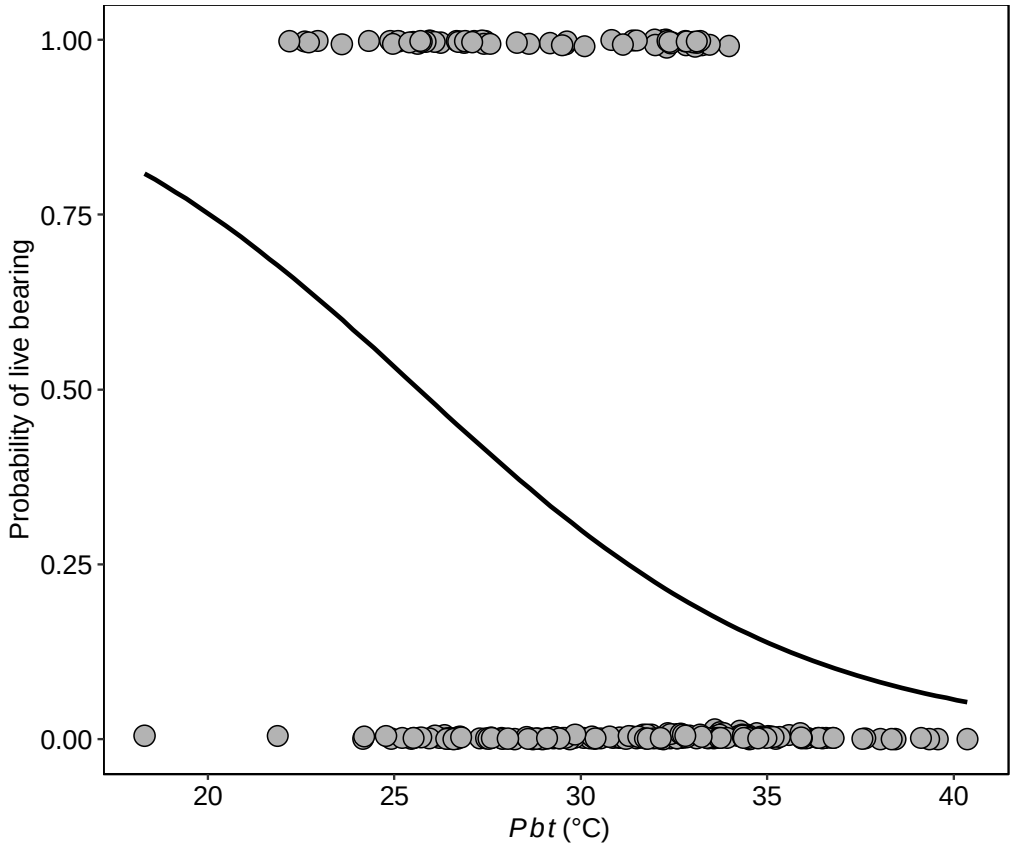

Figure 3. Probability of being a live-bearing species in relation to female non-gravid preferred body temperature (P_{bt}). The line and shading represent a logistic regression curve ± 95% confidence interval. P_{bt} was obtained for 103 oviparous and 61 viviparous species.

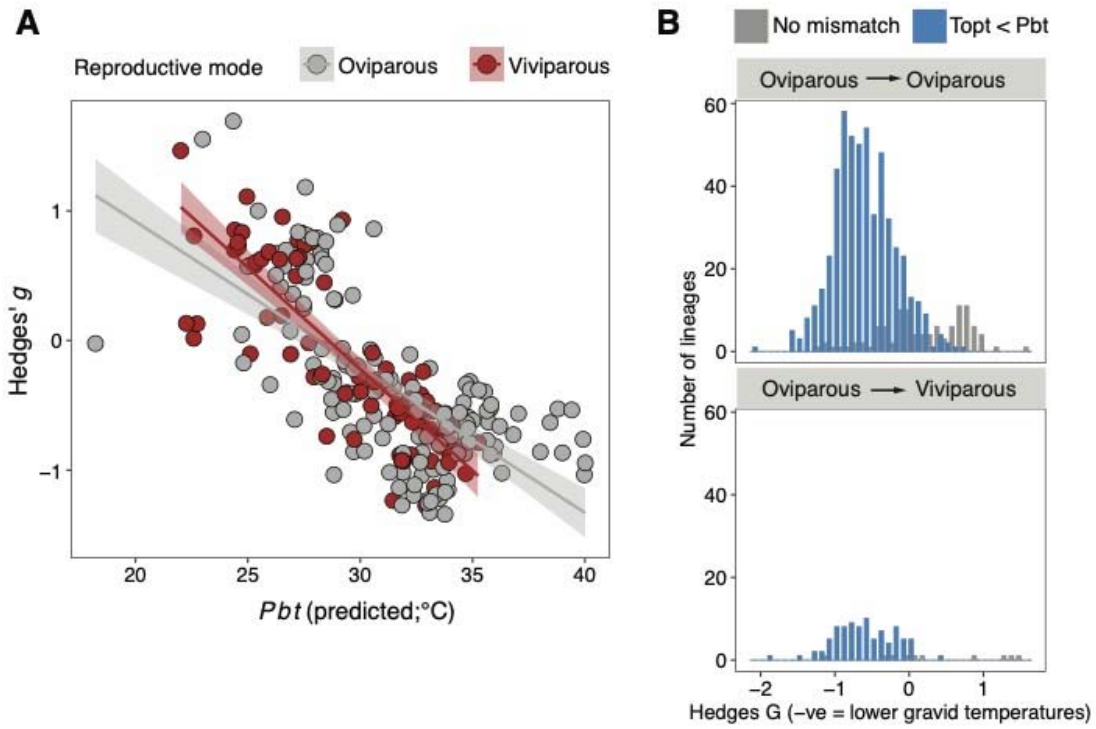

Figure 4. Female behavioural plasticity resolves the constraints imposed by thermal physiology, such that transitions to live-bearing (viviparity) occur despite mismatches between females and embryos. (A) Adjustment of body temperature in gravid females in relation to their non-gravid preferred body temperature (P_{bt}) across 52 extant oviparous and viviparous lizards and snakes (Hedges' g; $P_{bt\text{-}g}$ - $P_{bt\text{-}ng}$). Regression lines ± 95% credible intervals are plotted. High values of Hedges' g signify an increase in body temperature when gravid while low values imply a decrease. (B) The adjustment of body temperature by females in lineages that remained egg laying (top panel) and lineages that evolved live bearing (bottom panel) according to the presence or absence of a thermal mismatch between females and their offspring (inferred from Hedges' g data shown in panel A).