Behavioural adaptations in egg laying ancestors facilitate evolutionary transitions to live birth

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

#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 over 100 times in reptiles. One significant feature of this transition is the thermal conditions experienced by developing embryos. Egg-laying lizards and snakes often have preferred body temperatures that are lethal to developing embryos that 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 behaviourally 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[1–3]. 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[3,4]. By retaining embryos throughout development, mothers can buffer offspring from suboptimal nest temperatures, thereby ensuring higher hatching success and increased offspring viability[4]. Live bearing also allows females to maintain offspring at higher temperatures than those of potential nest sites, effectively increasing developmental rates to facilitate early-season hatching[5,6]. This transition has allowed reptile species to diversify and persist in cool climates globally[7].

Despite its adaptive advantage, how viviparity initially evolved is unclear. A major challenge to the evolution of viviparity is that egg-laying adult lizards and snakes typically have preferred body temperatures that exceed the upper lethal limit of embryos[8]. For example, the average nest temperature for the Iberian emerald lizard, *Lacerta schreiberi*, is 24 °C, rarely exceeding 30 °C, while the preferred body temperature of females is 33 °C[9]. Since embryos are well adapted to the temperatures they typically experience in the nest, they generally have limited capacity to sustain development at temperatures outside a narrow thermal window[10,11]. Prolonged exposure to temperatures beyond this range can result in offspring abnormalities and even death[12]. A mismatch between thermal optima of embryos and females should therefore prevent mothers from retaining embryos throughout development, inhibiting evolutionary transitions to live birth[8].

Live birth has nevertheless evolved over 100 times in squamate reptiles[13,14]. 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 oviparity in species 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 the mismatch 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. In addition, 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, now 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). 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[15].

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 (4162 species). Prior to running analyses, we estimated the phylogenetic signal in all traits to ensure it was possible to infer ancestral values of traits given data limitations. We found that there were high levels of phylogenetic signal in all traits (Tables S2, S6 & S8). We

further tested if models could accurately predict ancestral states using limited data by restricting data on reproductive mode to just those species that also had thermal data (224 species). These analyses showed that it was possible to predict the reproductive modes of 4162 species and their ancestral nodes with 94% accuracy from data on just these 224 species (Figure 3, Table S13).

Given phylogenetic models could accurately predict missing values and ancestral states we proceeded with the following analyses: First, the ancestral states of preferred body temperature of females and embryos were estimated to quantify the thermal conflict between mothers and embryos at the origins of viviparity. Second, we tested the prediction that transitions to viviparity have been restricted to oviparous lineages with a low degree of maternal-embryo mismatch. 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.

Results and Discussion

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

Across oviparous lizards and snakes, there were many species with a 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 = 4.42, 95% Credible Interval (CI): -2.98, 11.53, *pMCMC* = 0.12. Figure 1 & 2; Table S3). The female P_{bt} of oviparous species is on average 5 °C higher than 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[16]). P_{bt} and T_{opt} of oviparous species were both estimated to evolve slowly (Figure 1; MR-BPMM: P_{bt} phylogenetic signal (phylo H^2): 0.94, CI: 0.84, 0.99. T_{opt} phylo H^2: 0.94, CI: 0.78, 0.99. Table S3) and female P_{bt} and offspring T_{opt} only show weak evidence of coevolving (Figure 2; phylogenetic correlation (MR-BPMM): PM = 0.69, CI: -0.22, 0.90, *pMCMC* = 0.21. Table S4), 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 = -0.27, CI: -0.49, -0.01, *pMCMC* = 0.03; Figure 3; Table S6). This is consistent with the hypothesis that viviparity is an adaptation to cool climates[2,3,17]. 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 = 0.07, CI: -

1.80, 1.20, *pMCMC* = 0.41; Table S7). Additionally, live-bearing lineages with oviparous ancestors showed higher P_{bt} than those with live-bearing ancestors (MR-BPMM, PM = -1.46, CI: -2.79, 0.77, *pMCMC* = 0.18; Table S7).

Ancestral reconstructions of *Pbt* and *Topt* of oviparous species showed that transitions from egg laying to live bearing were no more likely to occur when adult and embryo thermal optima were similar (proportion of lineages with thermal mismatches: ancestors of live bearing species 69%, ancestors of egg laying species 78%. Table S5). Only eight of the estimated 26 evolutionary origins of live birth had adult and embryo thermal optima that overlapped, which was similar to the proportion of lineages where live bearing did not occur (Table S5). In the remaining 18 egg-laying ancestors of live-bearing species, the thermal optimum of embryos was significantly lower than the preferred body temperature of adults (there were no cases of T_{opt} being significantly higher than P_{bt}; Table S5). In summary, the degree of similarity between adult P_{bt} and embryo T_{opt} does not predict transitions towards viviparity, 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[18]. 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 those 52 species, there was a negative phylogenetic correlation between P_{bt} and the degree to which females reduced their body temperature when gravid (MR-BPMM: PM = -0.38, CI: -0.71, 0.13, *pMCMC* = 0.08. Figure 4; Table S8). Females with high *Pbt* have therefore evolved to behaviourally reduce their body temperature when gravid to a greater extent than females with low *Pbt*. This relationship was not restricted to viviparous lineages, as would have been expected if female behavioural plasticity is an adaptation that evolves in response to selection for embryo retention. Instead, comparable phylogenetic correlations between Hedges' *g* and *Pbt* were evident in both oviparous and viviparous taxa, showing that behavioural adjustment of temperature when gravid occurs to a similar extent in both egg-laying and live-bearing species (Egg laying: PM = -0.83, CI: -0.98, -0.02, *pMCMC* = 0.05. Live bearing: PM = -0.90, CI: -0.99, -0.25, *pMCMC* = 0.02. Table S9). Furthermore, when we reconstructed the values of Hedges' *g* (i.e., the effect of gravidity on body temperature) for ancestors of egg-laying and live-bearing species, we found that thermal mismatches between adults and embryos were resolved by females adjusting their body temperature when gravid (Figure 4A; MR-BPMM: Egg-laying Hedges' *g* PM = -0.71, CI = -1.21, -0.30, *pMCMC* < 0.01. Live bearing Hedges' *g* PM = -0.74, CI = -1.34, -0.31, *pMCMC* < 0.01. Table S10). As a result, Hedges' *g* did not differ between the ancestors of oviparous and viviparous species (Figure 4B; Table S9-S11), suggesting that behavioural plasticity was likely present prior to the emergence of live birth.

Our results illustrate 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 that paves the way for 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[19]). 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[8,10]. 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 temperature regulation put in place long before live birth emerged.

We suggest that this 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, 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[2,5,6,20,21].

In conclusion, female thermoregulatory behaviour is a key adaptation that overcomes the constraints imposed by thermal mismatches between mothers and embryos, and has facilitated both the evolution of live birth, and 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 collected existing data on reproductive mode (oviparous versus viviparous), the female preferred body temperature (P_{bt}) ("P_{bt} dataset"), the temperature at which hatching success was maximised (T_{opt}) ("T_{opt} dataset"), and 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.

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[22] 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. Given that T_{opt} had a strong phylogenetic signal (Phylogenetic Heritability (H^2) = 91.93%, 95% CI: 71.65 – 98.16%; Table S3) 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).

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[23].

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). The majority of 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) = -0.472, CI = -1.685, 0.557; PM (Viviparous) = 0.564, CI = -0.250, 1.313. Table S14. 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).* While for *S. equalis* all data were from an oviparous population, 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[24].

Quantification and Statistical Analysis

Prior to running our main analyses, we estimated the mode of evolution that best explained how traits evolved using Phylogenetic Generalised Least Squares (PGLS). Brownian motion was the best fitting model for all traits (Table S2). We then used Bayesian Phylogenetic Mixed Effects Models with single (BPMM) and multiple response variables (MR-BPMM), which enabled us to overcome the challenge of missing data across species, to estimate phylogenetic correlations between traits and reconstruct ancestral states. The robustness of ancestral reconstructions was examined using stochastic character mapping (SCM; Table S12). All analyses were conducted in R version 4.0.1[25].

Assessing modes of evolution using Phylogenetic Generalised Least Squares (PGLS)

An assumption of many phylogenetic comparative methods is that traits evolve via a Brownian motion process[26]. This assumption was assessed by estimating the mode of evolution that best describes the phylogenetic distribution of reproductive mode, *PBT* and

Topt using PGLS implemented with R packages 'ape' and 'nlme'[27,28]. This was done by comparing PGLS models with different correlation structures constructed from the phylogeny that corresponded to BM, Ornstein-Uhlenbeck (OU: corMartins in 'ape') and Accelerated Decelerated (ACDC: corMartins in 'ape') modes of evolution. In OU models' traits undergo a BM process but evolve towards a central tendency with a strength proportional to parameter alpha (α), and in accelerated/decelerated models of evolution (ACDC) traits diverge under a BM model, but rates of evolution accelerate or decelerate over time according to parameter g (Acceleration: $g < 1$. BM: $g = 1$. Deceleration $g > 1$). Estimating parameters, such as α and g, directly from the data can be unstable[29,30]. Therefore, we estimated values of α (OU) and Bloomberg's g (ACDC) by fixing α and g at 0.1 increments within the range of 0 to 10 and used AIC values to identify best-fit models. For BM models AIC values were taken from models with an optimised value of lambda rather than a strict BM model (lambda = 1). For all traits there was significant phylogenetic signature and BM best explained the way traits have evolved (Table S2: R code: Section 0).

Bayesian Phylogenetic Mixed Effects models (BPMM)

We implemented BPMMs in *R* with the *MCMCglmm* package[31]. All traits (T_{opt}, P_{bt}, *Hedges' g*) were modelled with Gaussian error distributions apart from the probability of viviparity which was modelled using a binary error distribution with a logit link function. The sample sizes for each species varied across traits, which was accounted for in models by weighting each data point by their inverse sampling variances using the 'mev' term in *MCMCglmm.* Samples sizes were not available for reproductive mode as this is measured without error due to it being invariant within populations. Therefore, the mev term for reproductive mode was specified as 0.

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 the 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.

Phylogenetic relationships were modelled using a published phylogeny comprising over 4000 species of squamates[32] that was pruned to the 224 species in our dataset using the *ape* package in *R*[27]. All analyses were conducted on the trimmed tree (Figure 1), and ancestral states were then verified using all available reproductive mode data (4162 species) and the full tree – all results remained the same (See verification analyses. Table S13, Figure S3). The random effect *animal*, with a covariance matrix derived from the phylogenetic tree was included in all models[31]. We calculated the phylogenetic signature (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[33]. More traditional methods were also used to validate estimates of phylogenetic signal. Pagel's lambda (λ) from *pgls* models[34,35] and Blomberg's K (*K*) in *phytools* using the function *phylosig*[36,37], which gave comparable estimates to BPMMs (see R code section 0).

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 allow missing data in response variables to be predicted, which was particularly useful as values of traits were not available for all species (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[15]. All traits had high phylogenetic signal (phylogenetic heritability >80%) producing high correspondence between raw and predicted values (Figure S2; See also Verification analyses "2. *Checking the ability of BPMMs to predict missing data"*). As a result, our BPMMs enabled us to deal with the fact that not all traits have been measured in all species.

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. 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[38]. 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. 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 total 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.

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; [31]) 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 [39]. 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.

Specific analyses

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 S3). **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 S4). **3)** 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 in the phylogenetic tree using a BPMM with the probability of live bearing as the response variable (R code model M3.0). 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 S5). Separate models were used for reproductive mode and P_{bt}, T_{opt} and Hedges' *g* to ensure that ancestral estimates were not influenced by the strong correlation between live bearing and P_{bt} in extant species. **4)** 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). **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 & S10-S11). **6)** The phylogenetic correlation between P_{bt} and Hedges' *g* was estimated across all species (R code model M3.1. Table S8), 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 S9).

Verification analyses

1. Checking ancestral reconstructions with Stochastic character mapping (SCM)

The robustness of ancestral state reconstructions of reproductive mode was examined by assessing the correspondence between the estimates from MCMCglmm with those obtained from SCM. The SCM approach calculates the conditional likelihood that each ancestral node is oviparous or viviparous based on an estimated transition rate matrix (Q) between reproductive modes and the length of the branch associated with that node. Based on these conditional likelihoods, ancestral states at each node are stochastically simulated and used in combination with observations at the tips to reconstruct a character history along each branch. Each character history is simulated using a continuous-time Markov model where changes between states and the time spent in each state is modelled as a Poisson process (see [40] for details). We found extremely good correspondence between estimates of ancestral states obtained from MCMCglmm models and SCM showing inferences of ancestral states were robust to methodological details (Table S12).

2. Checking the ability of BPMMs to predict missing data

We examined how well models predicted missing values by running a BPMM of reproductive mode on the full tree (4162 species) using only phenotypic data from the 224 species included in the main analyses (R code model M5.3: Table S13). The model settings were identical to those described above for modelling reproductive mode. The accuracy of the model predictions for tips in the tree were checked against raw data[41], and model predictions for tree nodes were checked against reconstructions of ancestral states obtained from a BPMM using the full tree and full dataset (R code model M5.2. Table S13).

3. 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 S14).

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., C.K.C. and T.U.; 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., C.K.C. and T.U.; Writing – original draft, A.K.P, N.F., C.K.C and T.U.; Writing – Reviewing & Editing, A.K.P., N.F., D.W.A.N., G.M.W., C.K.C. and T.U.; Visualization, A.K.P., N.F., D.W.A.N. and C.K.C.; Supervision, C.K.C. and T.U.; Funding acquisition, A.K.P. and T.U.

Competing Interest Statement: The authors declare no competing interests.

References

1. Doody, J. S. & Moore, J. A. Conceptual model for thermal limits on the distribution of reptiles. *Herpetological Conservation and Biology* **5**, 283–289 (2010).
2. Shine, R. The evolution of viviparity in reptiles: An ecological analysis. in *Biology of the Reptilia* vol. 15 605–694 (John Wiley & Sons, Ltd, 1985).
3. Lambert, S. M. & Wiens, J. J. Evolution of viviparity: a phylogenetic test of the cold-climate hypothesis in Phrynosomatid lizards. *Evolution* **67**, 2614–2630 (2013).
4. Shine, R. A new hypothesis for the evolution of viviparity in reptiles. *The American Naturalist* **145**, 809–823 (1995).
5. Warner, D. A. & Shine, R. Fitness of juvenile lizards depends on seasonal timing of hatching, not offspring body size. *Oecologia* **154**, 65–73 (2007).
6. Le Henanff, M., Meylan, S. & Lourdais, O. The sooner the better: reproductive phenology drives ontogenetic trajectories in a temperate squamate (Podarcis muralis). *Biol J Linn Soc* **108**, 384–395 (2013).
7. Ma, L., Buckley, L. B., Huey, R. B. & Du, W.-G. A global test of the cold-climate hypothesis for the evolution of viviparity of squamate reptiles. *Global Ecology and Biogeography* **27**, 679–689 (2018).
8. Beuchat, C. A. Temperature effects during gestation in a viviparous lizard. *Journal of Thermal Biology* **13**, 135–142 (1988).
9. Monasterio, C., Shoo, L. P., Salvador, A., Iraeta, P. & Díaz, J. A. 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 (2013).
10. Sanger, T. J., Kyrkos, J., Lachance, D. J., Czesny, B. & Stroud, J. T. 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 (2018).

11. Du, W.-G., Shine, R., Ma, L. & Sun, B.-J. 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 (2019).
12. Andrews, R. M., Mathies, T. & Warner, D. A. Effect of Incubation Temperature on Morphology, Growth, and Survival of Juvenile Sceloporus undulatus. *Herpetological Monographs* **14**, 420–431 (2000).
13. Blackburn, D. G. Squamate Reptiles as Model Organisms for the Evolution of Viviparity. *Herpetological Monographs* **20**, 131–146 (2006).
14. Blackburn, D. G. Evolution of viviparity in squamate reptiles: Reversibility reconsidered. *J Exp Zool B Mol Dev Evol* **324**, 473–486 (2015).
15. Molina-Venegas, R. *et al.* Assessing among-lineage variability in phylogenetic imputation of functional trait datasets. *Ecography* **41**, 1740–1749 (2018).
16. Noble, D. W. A., Stenhouse, V. & Schwanz, L. E. Developmental temperatures and phenotypic plasticity in reptiles: a systematic review and meta-analysis. *Biological Reviews* **93**, 72–97 (2018).
17. Watson, C. M., Makowsky, R. & Bagley, J. C. 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 (2014).
18. Beuchat, C. A. Reproductive Influences on the Thermoregulatory Behavior of a Live-Bearing Lizard. *Copeia* **1986**, 971–979 (1986).
19. Andrews, R. M. & Mathies, T. Natural History of Reptilian Development: Constraints on the Evolution of Viviparity. *BioScience* **50**, 227–238 (2000).

20. Andrews, R. M. Evolution of viviparity in squamate reptiles (Sceloporus spp.): a variant of the cold-climate model. *Journal of Zoology* **250**, 243–253 (2000).
21. Uller, T. & Olsson, M. Offspring size and timing of hatching determine survival and reproductive output in a lizard. *Oecologia* **162**, 663–671 (2010).
22. Noble, D. W. A. *et al.* A comprehensive database of thermal developmental plasticity in reptiles. *Sci Data* **5**, 1–7 (2018).
23. Borenstein, M., Hedges, L. V., Higgins, J. P. & Rothstein, H. R. Effect Sizes Based on Means. in *Introduction to Meta-Analysis* 21–32 (John Wiley & Sons, Ltd, 2009). doi:10.1002/9780470743386.ch4.
24. Recknagel, H., Kamenos, N. A. & Elmer, K. R. 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 (2018).
25. R Core Team. *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* (2020).
26. *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice*. (Springer-Verlag, 2014). doi:10.1007/978-3-662-43550-2.
27. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
28. Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & Team, R. C. nlme: Linear and nonlinear mixed effects models. *R package version* **3**, 111 (2013).
29. Paradis, E. *Analysis of Phylogenetics and Evolution with R.* (Springer-Verlag, 2012). doi:10.1007/978-1-4614-1743-9.
30. Hansen, T. F. Stabilizing selection and the comparative analysis of adaptation. *Evolution* **51**, 1341–1351 (1997).

31. Hadfield, J. D. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software* **33**, 1–22 (2010).
32. Zheng, Y. & Wiens, J. J. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics and Evolution* **94**, 537–547 (2016).
33. Nakagawa, S., Johnson, P. C. D. & Schielzeth, H. 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 (2017).
34. Orme, D. *et al*. CAPER: comparative analyses of phylogenetics and evolution in R. *Methods in Ecology and Evolution* **3**, 145–151 (2013).
35. Freckleton, R. P., Harvey, P. H. & Pagel, M. Phylogenetic analysis and comparative data: a test and review of evidence. *Am. Nat.* **160**, 712–726 (2002).
36. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**, 217–223 (2012).
37. Blomberg, S. P., Garland, T. & Ives, A. R. Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits Are More Labile. *Evolution* **57**, 717–745 (2003).
38. Buuren, S. van & Groothuis-Oudshoorn, K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* **45**, 1–67 (2011).
39. Brooks, S. P. & Gelman, A. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics* **7**, 434–455 (1998).
40. Bollback, J. P. SIMMAP: Stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* **7**, 88 (2006).
41. Halliwell, B., Uller, T., Holland, B. R. & While, G. M. Live bearing promotes the evolution of sociality in reptiles. *Nature Communications* **8**, 2030 (2017).

Figures

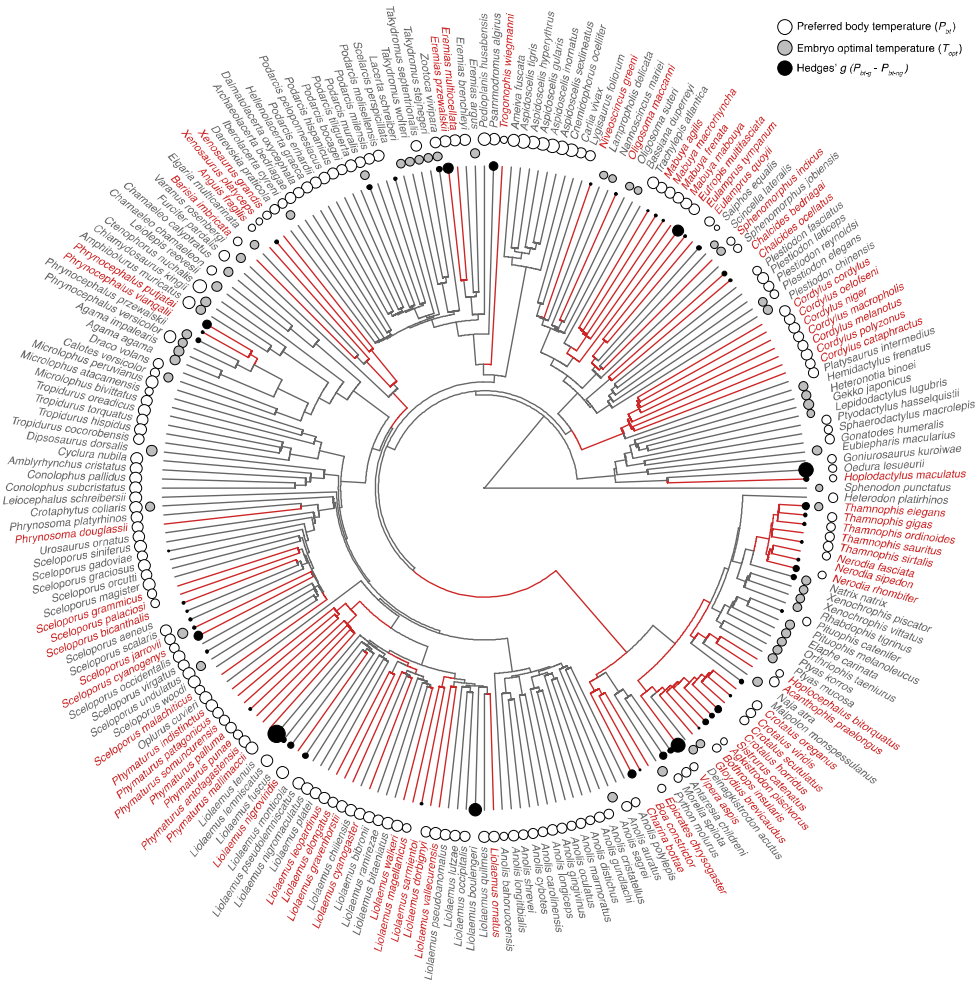

Figure 1. Variation in reproductive mode, preferred body temperature and plasticity in female thermo-regulatory behaviour across 224 species of squamate reptile (lizards and snakes). Tips and branches are coloured according to reproductive modes (red = live bearing/viviparous, grey = egg laying/oviparous; branch colours represent predicted ancestral values; Table S10). 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).

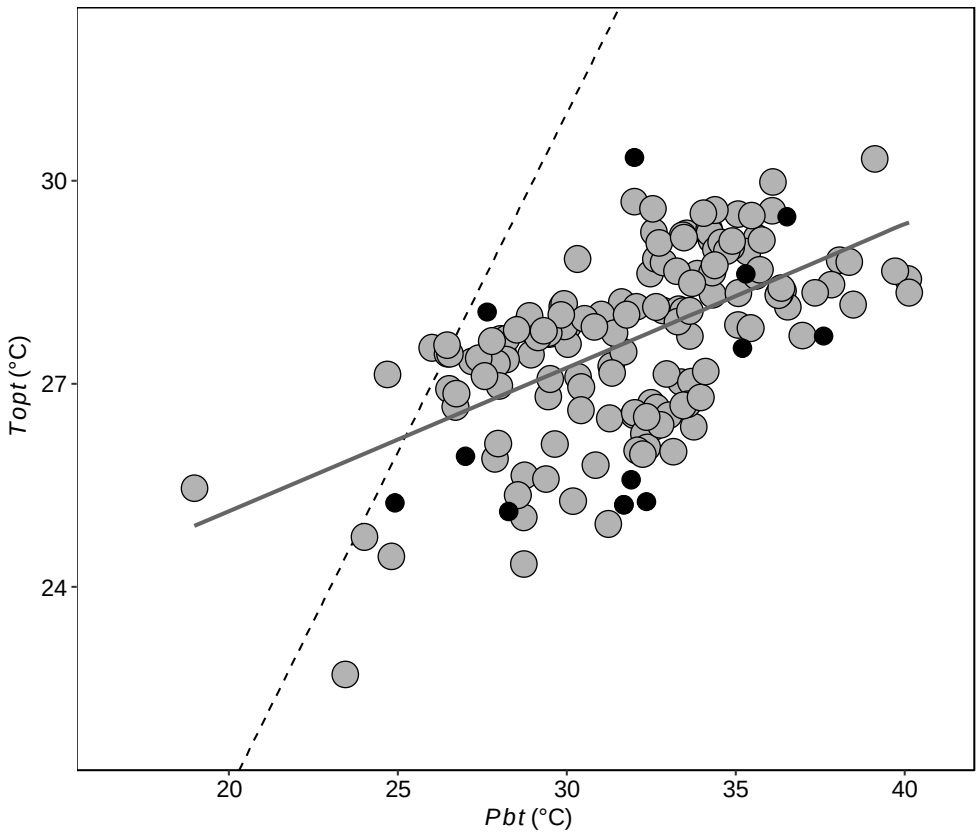

Figure 2. Mismatches in the thermal optima of mothers and embryos across 147 oviparous species of lizards and snakes. The relationship between female preferred body temperature (P_{bt}) and the temperature that maximizes hatching success of embryos (T_{opt}) predicted by a multi-response Bayesian Phylogenetic Mixed Model across egg laying species (grey data points, regression lines ± 95% confidence intervals are plotted). Raw data for 12 species with both *Pbt* and *Topt* 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 S4). 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 S2).

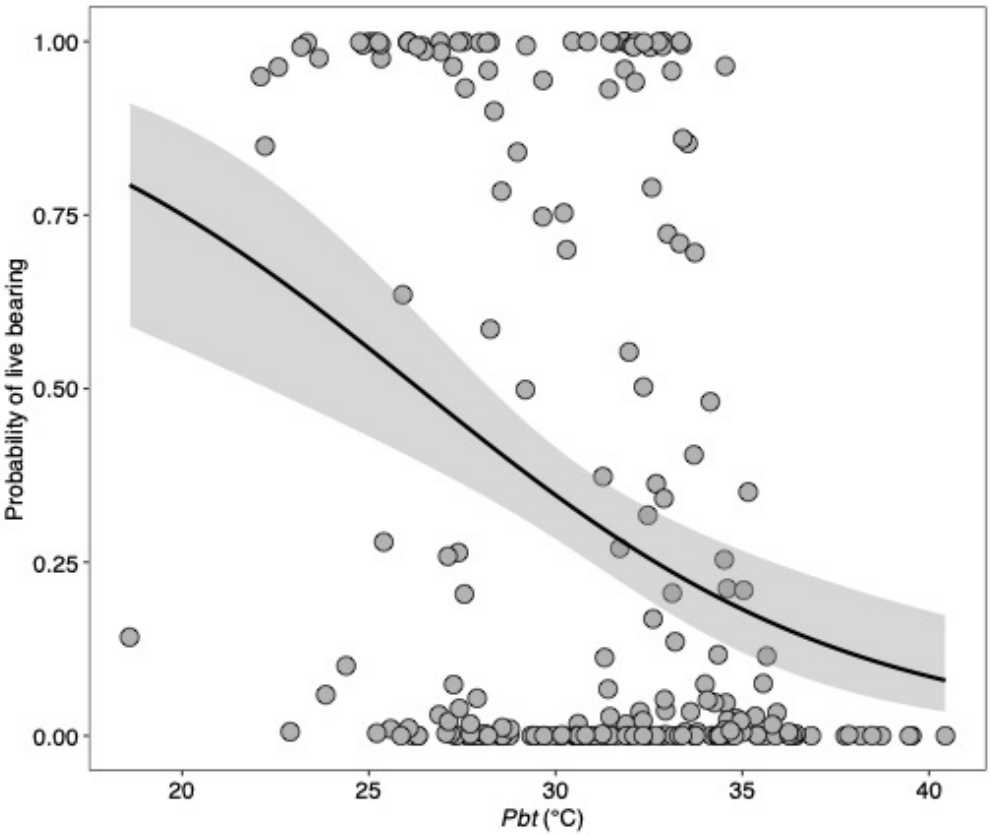

Figure 3. Probability of being a live-bearing species in relation to female non-gravid preferred body temperature (P_{bt}). The line and shading represent a logistic regression curve ± 95% confidence interval. P_{bt} was obtained for 103 oviparous and 61 viviparous species.

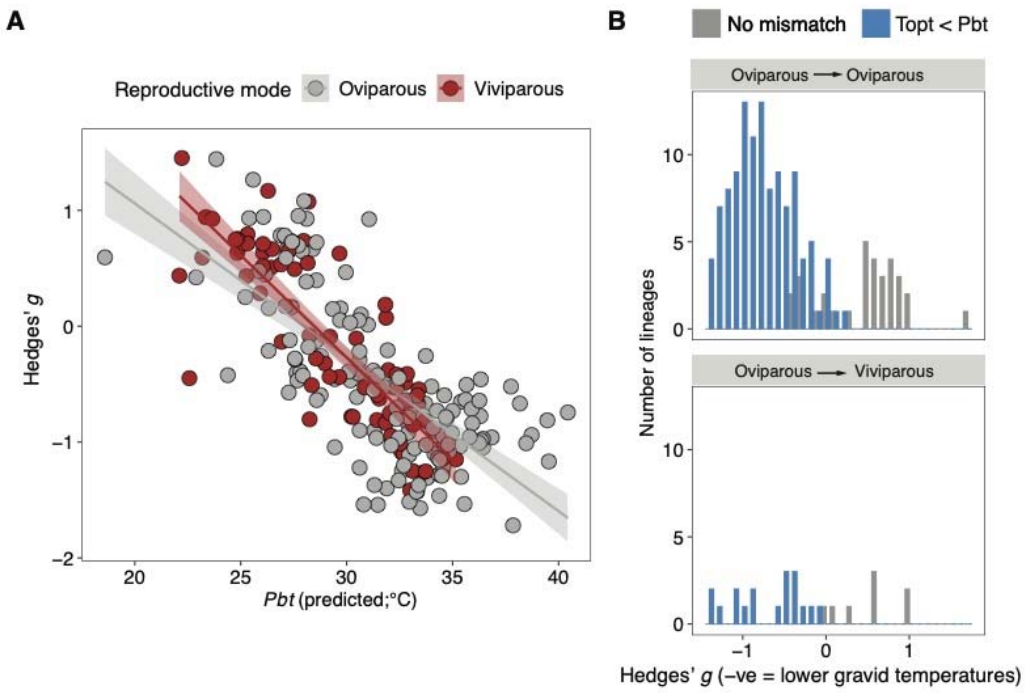

Figure 4. Female behavioural plasticity resolves the constraints imposed by thermal physiology, such that transitions to live-bearing (viviparity) occur despite mismatches between females and embryos. (A) Adjustment of body temperature in gravid females in relation to their non-gravid preferred body 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).