RESEARCH ARTICLE

Body temperature is a repeatable trait in a free-ranging passerine bird

Simon Tapper[1],*, Joseph J. Nocera[2] and Gary Burness[3]

ABSTRACT

Body temperature (T_b) affects animal function through its influence on rates of biochemical and biophysical reactions, the molecular structures of proteins and tissues, and, ultimately, organismal performance. Despite its importance in driving physiological processes, there are few data on how much variation in T_b exists within populations of organisms, and whether this variation consistently differs among individuals over time (i.e. repeatability of a trait). Here, using thermal radio-frequency identification implants, we quantified the repeatability of T_b, both in the context of a fixed average environment (~21°C) and across ambient temperatures (6–31°C), in a free-living population of tree swallows (*Tachycineta bicolor*, n=16). By experimentally trimming the ventral plumage of a subset of female swallows (n=8), we also asked whether the repeatability of T_b is influenced by the capacity to dissipate body heat. We found that both female and male tree swallow T_b was repeatable at 21°C (R=0.89–92), but female T_b was less repeatable than male T_b across ambient temperature (R_{female}=0.10, R_{male}=0.58), which may be due to differences in parental investment. Trimmed birds had on average lower T_b than control birds (by ~0.5°C), but the repeatability of female T_b did not differ as a function of heat dissipation capacity. This suggests that trimmed individuals adjusted their T_b to account for the effects of heat loss on T_b. Our study provides a first critical step toward understanding whether T_b is responsive to natural selection, and for predicting how animal populations will respond to climatic warming.

KEY WORDS: Heat dissipation capacity, Phenotypic flexibility, Reaction norm, Sex differences, Tree swallow

INTRODUCTION

Reaction norms can be used to simultaneously quantify both the level of phenotypic flexibility across an environmental gradient (i.e. reaction norm slope) and the phenotypic expression of a trait in the population-average environment (i.e. reaction norm intercept). Reaction norm components of several flexible physiological traits have been quantified, many of which are related to energy metabolism (e.g. basal metabolic rate, glucocorticoids: reviewed in Biro and Stamps, 2010) and immunity (e.g. Schreier and Grindstaff, 2020; Tieleman et al., 2010). However, in endotherms, individual variation in the phenotypic flexibility of body

temperature (T_b) has been understudied by evolutionary physiologists, despite its importance in shaping organismal performance (Angilletta et al., 2010). T_b exhibits substantial plasticity in endotherms (Boyles et al., 2013), as evidenced by adjustments in T_b in response to different environmental challenges (e.g. food availability, thermal stress), activity level, daily fluctuations in rhythmicity (e.g. circadian and ultradian rhythms) and regional heterothermy within individuals as a consequence of heat flow across the body (Angilletta et al., 2010). Yet, studies investigating variation in T_b (and other thermoregulatory traits) have primarily focused on differences among species and populations (e.g. Noakes and McKechnie, 2019; Noakes and McKechnie, 2020; Smit et al., 2013), or among groups within populations (e.g. age and sex: Cai et al., 2016; Gagnon and Kenny, 2012; Andreasson et al., 2020a,b) rather than among individuals within a single population (e.g. Pessato et al., 2020; Bozinovic, 2007; Møller, 2010).

Evidence from laboratory studies, both on wild-caught and lab-raised individuals, indicates that among-individual variation in T_b exists at the individual level, and may be related to differences in metabolic traits that vary across environments (Briga and Verhulst, 2017). For instance, in birds, individuals with higher basal metabolic rate (a trait with reported genetic basis: Rønning, et al., 2007) also maintain higher T_b during exposure to temperatures below thermoneutrality (Briga and Verhulst, 2017; Stager et al., 2020), and in mice, individual consistency in T_b variation predicts torpor use (Nespolo et al., 2003; Boratyński et al., 2019). Further, variation among individuals in resting and basal metabolic rate has been widely reported in laboratory and free-ranging studies (e.g. Boratyński et al., 2017; Broggi et al., 2009; Hayes et al., 1998; Nespolo and Franco, 2007; Rønning et al., 2005; White et al., 2013). Finally, genetic variation in T_b has been reported in both domestic and laboratory species, including chickens (h^2 range, 0.10–0.20: Li et al., 2020; van Goor et al., 2015), dairy cattle (h^2=0.17: Dikmen et al., 2012), sows (h^2=0.35: Gourdine et al., 2017) and mice (h^2=0.36: Connolly and Lynch, 1981).

While there is evidence demonstrating consistent individual differences in thermoregulation in laboratory and domestic animals, far fewer studies have formally examined among-individual variation in T_b in free-ranging species (e.g. Bozinovic, 2007; Dammhahn et al., 2017; Møller, 2010; Nespolo et al., 2003, 2010; Pessato et al., 2020). Free-living animals often experience multiple environmental challenges simultaneously, and the degree of individual consistency and plasticity in T_b might be adjusted differently depending on the environmental context, physiological condition and life-history strategy of the individual (e.g. Cornelius et al., 2017). For example, metabolic rate in free-living animals has been reported to be less repeatable than in laboratory animals, which may be a consequence of how animals differ in their response to the same conditions over time (Auer et al., 2016). Further, exposure to certain environmental conditions during development, such as high environmental temperatures, may induce epigenetic changes in

[1]Environmental and Life Sciences Graduate Department, Trent University, 1600 West Bank Drive, Peterborough, ON, Canada, K9L 0G2. [2]Faculty of Forestry and Environmental Management, University of New Brunswick, 28 Dineen Drive, Fredericton, NB, Canada, E3B 5A3. [3]Department of Biology, Trent University, 1600 West Bank Drive, Peterborough, ON, Canada, K9L 0G2.

*Author for correspondence (simontapper@trentu.ca)

S.T., 0000-0003-3237-291X; J.J.N., 0000-0002-5913-8851; G.B., 0000-0002-1695-7179

aspects of thermoregulatory physiology (see Ruuskanen et al., 2021, for a review), such as heat shock protein regulation (Kisliouk et al., 2017) and morphology (Burness et al., 2013; see Nord and Giroud, 2020, for a review). Given, that wild animals experience fluctuating environmental conditions both intra- and inter-annually, there may be greater variation in the magnitude and type of epigenetic and parental effects experienced by wild versus laboratory animals.

To understand the degree to which T_b consistently varies among free-living birds, we examined the repeatability of T_b in tree swallows, *Tachycineta bicolor*, across differing ambient temperatures (T_a). We asked: (1) do birds consistently differ from each other in their average T_b (i.e. is T_b repeatable at the reaction norm intercept)?; and (2) do birds consistently differ from each other in their T_b response across ambient temperature T_a (i.e. is the reaction norm slope repeatable)? Here, we used previously published data from our research group (Tapper et al., 2020a,b) to examine repeatability in swallows.

Because inter-specific variation exists among thermoregulatory traits (Pessato et al., 2020; Stager et al., 2020), we predicted that in both female and male swallows, T_b would be repeatable, both in terms of the average trait expression (i.e. intercept) and how individuals vary T_b across T_a (i.e. slope). However, because female tree swallows have a brood patch, whereas males do not, there may be differences in heat dissipation capacity between the sexes (Nord and Nilsson, 2019). Therefore, we may expect males to forage closer to their maximally allowable T_b relative to females, which could lead to differences in repeatability between the sexes. This could occur if, for example, the upper threshold on T_b does not vary between individuals for biochemical reasons. Alternatively, if both sexes reduce their activity level at high temperatures, meaning that hyperthermia may be a limiting factor, there may be similar repeatability in T_b between male and female swallows.

In our previous studies (Tapper et al., 2020a,b), we experimentally removed the ventral feathers from a sample of breeding females to increase heat dissipation capacity (i.e. trimmed treatment). Although testing for a treatment effect on T_b repeatability was not an explicit goal of this study, previously we found that control birds had higher T_b than trimmed birds (by ~0.5°C) across temperature (6–31°C) (Tapper et al., 2020b), suggesting that differences in heat dissipation capacity could lead to differences in the repeatability of T_b. Following the same reasoning as above, control and trimmed birds may differ in T_b repeatability if control birds forage relatively closer to their maximally allowable T_b, and if the upper threshold on T_b does not vary between individuals. Alternatively, T_b repeatability may be similar between experimental treatments if individuals tend to adjust their activity level to avoid overheating (Tapper et al., 2020b). In this case, the effects of heat dissipation capacity on T_b should be outweighed by behavioural (or physiological) differences among individuals that directly relate to individual control of T_b.

MATERIALS AND METHODS
Field methods
Our dataset is derived from Tapper et al. (2020b), in which a detailed description of our study sites, field methods and reproductive parameter definitions can be found. However, given that environmental variables such as wind speed, relative humidity and precipitation could affect individual variation in thermoregulation, we present these summary statistics for our study sites here (Table 1). We performed this study in May–July 2018, on two nest-box breeding populations of tree swallows, *Tachycineta bicolor* (Vieillot 1808), in Peterborough, ON, Canada. We captured female tree swallows during incubation (day 7–10, $n=16$) and, when possible, their respective male partners during early nestling provisioning (day 4–6, $n=10$). Upon capture, we implanted individuals with a thermal-sensitive passive integrated transponder (PIT) tag (accuracy ±0.5°C, Bio-Therm13, Biomark, Boise, IA, USA) subcutaneously into the nape of the neck to remotely monitor T_b. PIT tags can shift position post-implantation and if such shifts occurred in our study, this could change among-individual differences in T_b (i.e. increase or decrease repeatability estimates). This will depend on whether the tag was measuring T_b from different body regions within the same individual at different times. However, several studies have implanted PIT tags in the nape and reported minimal issues with respect to tag movement (e.g. Nord et al., 2013, 2016; Oswald et al., 2018), and thus variation in subcutaneous T_b from PIT tag placement seems unlikely. Similarly, in the unlikely event that a tag was placed too close to the skin surface, estimates of T_b could be underestimated as a result of the influence of environmental temperature. Finally, repeatability could be affected by differences in individual tag accuracy, but these differences are likely to be small, as reported in other studies (e.g. 0.2°C: O'Connor et al., 2021). We read PIT tags with Biomark HPR Plus readers (with the minimum time between consecutive reads of the same tag set to 10 s: Tapper et al., 2020b), which we connected to a loop antenna and positioned so that they encircled the nest box entrance.

We recaptured individual female swallows during early nestling provisioning (range: day 1–2 post-hatch) and performed our experimental manipulation (experimental trimming versus handling) to allow for increased heat loss. We assigned females to each treatment by flipping a coin between 'pairs' of birds, such that the first female captured received one treatment, while the second female captured received the alternative treatment. We repeated this process for each new dyad. In the case of individuals that had been experimentally manipulated in the previous year, they were typically assigned to the opposite treatment from that which they had experienced previously. Therefore, while we assigned pairs of females to each treatment based on a coin flip, this was not always possible, and on occasion we had to selectively alternate between treatment types. For each female, we either trimmed the contour and downy feathers covering the brood patch to expose the bare skin underneath (~7% of the surface area of the bird; trimmed treatment)

Table 1. Summary statistics for environmental variables that could affect heat transfer and individual variation in body temperature during the study period (May–July 2018)

	T_a (°C)	Wind speed (km h^{-1})	Total precipitation (mm)	Relative humidity (%)
Mean	19.4	6.3	2.0	69.2
Range (min., max.)	5.1, 34.6	0, 20.0	0, 28.5	24, 99
s.d.	6.2	4.0	5.4	21.6

The window of time included stretches from the first hatch day observed (30 May) to the last day of the data recording period (11 July, which was 14 days post-hatch for the last active nest). T_a, ambient temperature.

or ascribed them a control condition where they were handled but released with their feathers intact. All research was approved by the Trent University Animal Care Committee, in accordance with the Canadian Council on Animal Care (AUP no. 24747).

Data compilation, organization and statistical analyses

For data compilation and statistical analyses, we used R (version 4.0, R Core Team 2020; http://www.R-project.org/). To maximize our sample size (i.e. assess the repeatability of T_b under a broad range of T_a), we included all T_b measurements made between 2 and 16 days post-hatch, and between 05:00 h and 21:00 h (when swallows are diurnally active). We attempted to record individuals for 24 h at least 3 times throughout the nestling period, once between the periods of day 2–5, day 6–10 and day 11–16. While most individuals had balanced data across the nestling period (Fig. S1A), the T_b data of some individuals covered a wider range of T_a (6–31°C) than others (Fig. S1B). To control for differences in the number of T_b measurements among birds, and to minimize bias in our estimates of individual T_b variance, we calculated the weighted average hourly T_b (°C) for each bird. To do this, we used maximum likelihood to approximate the distribution of the number of reads within a single nest visit (gamma distribution, shape=0.56, scale=37.58) and the number of visits within an hour (normalized between 0 and 1, beta distribution, α=1.02, β=4.63). Approximating the distributions helped to smooth the probabilities and therefore avoid giving a certain number of reads (within a visit) or visits (within an hour) too much or too little weight. The mean number of reads per bird across females and males was 4984 (range 1559–15,917) and 1247 (range 393–3427), respectively. The mean number of visits per bird across females and males was 626 (range 198–1855) and 544 (range 193–1460), respectively. As in Tapper et al. (2020b), we excluded data from one bird that had abnormally low T_b (~2°C) between 05:00 h and 13:00 h on one day relative to its T_b between 05:00 h and 13:00 h on all other days (i.e. 13% of all observations for that bird).

To test whether T_b was repeatable in our tree swallows, we used Bayesian hierarchical mixed effects models, constructed using the package 'brms' (version 2.14, Bürkner, 2017). First, we quantified the repeatability and phenotypic flexibility of T_b of each sex, based on control bird data only. While removing the trimmed birds reduced our sample size from 29 ($n_{female}=16$, $n_{male}=13$) to 16 individuals ($n_{female}=8$, $n_{male}=8$), any differences in variance between treatments may have been confounded with sex differences on T_b if trimmed birds were included in the analysis. Thus, removing trimmed individuals increases the interpretability of our results.

In our model, we used a 'heterogeneous error structure', partitioning the different variance components (i.e. intercept, slope, intercept–slope co-variance, and residual error terms), into female ($n=8$) and male ($n=8$) groups, respectively. Because estimates of variance in reaction norm components (i.e. intercept and slope) differ depending on whether potential confounding factors are controlled for in the analysis, we included several covariates (see below) in our model to calculate the adjusted repeatability (Nakagawa and Schielzeth, 2010). We also mean-centred and standardized all covariates to two standard deviation units (Gelman, 2008). Our response variable was therefore mean hourly T_b, and our fixed predictors were treatment, T_a, lay date, brood size, nestling provisioning rate, and a 'sex×T_a' interaction. We note that in our preceding study on the effects of treatment on T_b (Tapper et al., 2020b), we modelled T_b as a polynomial function of T_a. Here, we chose to model T_b as a linear function of T_a to increase

interpretability of the reaction norm components (both intercepts and slopes) and to avoid overfitting our models (as we had insufficient data to properly model non-linear relationships per bird). We also ran our models with wind speed and relative humidity included as fixed effects (we did not have hourly precipitation data), but the results, and consequently our conclusions, remained similar to the model with these factors excluded. Therefore, we excluded these variables from the final analysis to reduce the number of model terms in our analysis. To determine the degree of phenotypic flexibility in T_b in response to changes in T_a, we also included random slopes for individual identity by T_a. To compare the degree of phenotypic flexibility in T_b between treatments, we ran a Levene's test on the slope estimates from the model. Lastly, we interpreted differences between sexes in the mean repeatability intercept ($R_{intercept}$) and slope (R_{slope}) using Bayes factors (K), as derived from one-way hypothesis tests (calculated with the Savage–Dickey density ratio method) in 'brms'. To keep our tests conservative, we asked whether the sexes differed from each other in $R_{intercept}$ and R_{slope} by a repeatability of >0.1, or 10% (e.g. $R_{intercept\ trimmed}-R_{intercept\ control}>0.1$), meaning that we considered a difference in $R≤0.1$ as practically equivalent to 0 (see Kruschke, 2018, for defining regions of practical equivalence). Support for a difference in repeatability would be given by $K>1$, with larger K values indicating stronger support and smaller K values weaker support (see Jeffreys, 1961, and Lee and Wagenmakers, 2013, for a guideline on interpretation of Bayes factors). While the choice of 0.1 is arbitrary, repeatability estimates from other avian physiological traits (e.g. basal metabolic rate, glucocorticoid levels) have been reported to cluster between 0.1 and 0.4 (Auer et al., 2016; Schoenemann and Bonier, 2018), and we therefore based our boundary on the low end of the repeatability estimates reported in the literature. Our priors for the intercept repeatabilities were beta distributed (α=12, β=3) for both sexes, and for female and male slope repeatabilities they were beta distributed (female, α=1, β=4; male, α=4, β=3).

Next, we quantified the repeatability and phenotypic flexibility of T_b in females ($n_{control}=8$, $n_{trimmed}=8$) to determine whether heat dissipation capacity (i.e. experimental trimming treatment) affected among-individual variation in T_b. Our second model was structured similarly to our first, except that population-level effects of sex and sex×T_a were replaced with the population-level effects of treatment and treatment×T_a. Additionally, instead of partitioning variance into female and male groups, we partitioned the variance components into control and trimmed groups, respectively. As above, we compared the phenotypic flexibility in T_b between treatments by running a Levene's test on the slope estimates from the model. We also tested for differences in the mean repeatability intercept and slope, with a statistical difference between groups in repeatability as >0.1. For both treatments, priors were beta distributed for the intercept (α=12, β=3) and slope (α=1, β=4).

For both the 'sex' and 'treatment' model, we used informative prior distributions for the population intercept (i.e. T_b, °C), T_a and feeding rate coefficients, which were based on model estimates in Tapper et al. (2020b). We used gamma distributions for the population intercepts (shape=84, scale=2; producing a peak density of ~42) and T_a coefficients (shape=1.2, scale=0.75; producing a peak density of ~0.3) to fix T_b, T_a and feeding rate estimates above 0. For feeding rate and treatment, we used normal distributions (mean=0 and s.d.=1) and for the random effect terms, we used gamma distributions (shape=1.75, scale=0.75; yielding a peak density of ~0.5). For the remaining model parameters (i.e. population-level fixed effects and residual error terms), we used
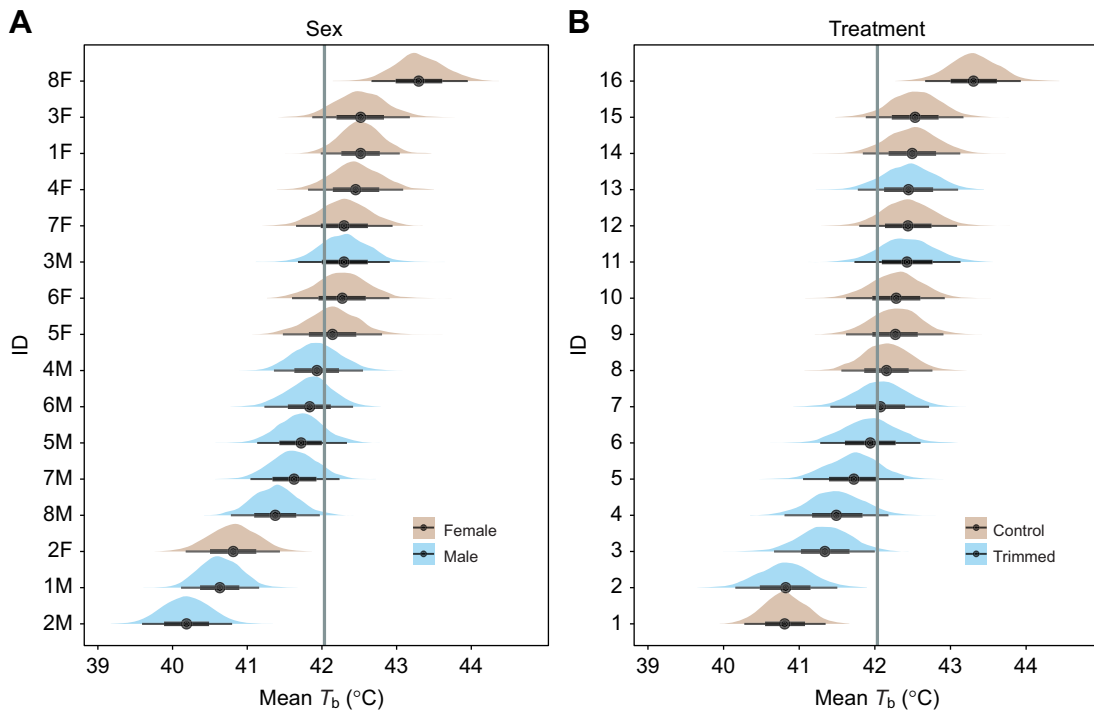
Fig. 1. Repeatability of subcutaneous body temperature (T_b) of tree swallows at an ambient temperature (T_a) of 21°C according to sex and treatment.
(A) T_b of (control) female and male provisioning tree swallows (n_{female}=8, n_{male}=8). (B) T_b of control and trimmed provisioning female tree swallows ($n_{control}$=8, $n_{trimmed}$=8). Repeatability estimates were predicted from a Bayesian hierarchical model. Histograms and black dots represent the posterior probability distribution and mean T_b estimates per individual, respectively. The black lines extending underneath the distributions represent 66% (thick black line) and 95% (thin black line) intervals. In A, pairs from the nest are identified by their ID number (e.g. 1F and 1M belong to the same nest). The solid grey line represents the mean T_b across all individuals.

Table 2. Differences between sexes in daytime subcutaneous body temperature (T_b) of control female tree swallows: fixed effects

Term	Estimate	95% CI [LCI, UCI]	ESS
Intercept	42.45	41.74, 43.30	2700
Sex	−0.85	−1.77, 0.05	2653
T_a	0.38	0.24, 0.51	2603
Brood size	−0.09	−1.33, 0.96	2689
Lay date	0.56	−0.55, 1.55	2684
Feeding rate	0.11	0.07, 0.15	2684
Sex×T_a	−0.03	−0.30, 0.24	2650

Parameter estimates (±95% credible interval, CI) for the population-level terms (i.e. fixed effects). n_{female}=8, n_{male}=8. Note that repeatability estimates are 'adjusted' (Nakagawa and Schielzeth, 2010) to control for the influence of population-level predictors (e.g. T_a) on mean T_b. LCI, lower confidence interval, UCI, upper confidence interval; ESS, effective sample size.

the 'brms' default non-informative priors, which follow Student's *t* distributions (with degrees of freedom=3, location=0, scale factor=2.5). We corrected for autocorrelation between adjacent T_b measurements by adding an AR1 correlation structure (ρ=0.64). We ran models for 10,000 iterations, discarding the first 1000 iterations as burn-in, and sampling every 10 iterations. $\hat{R}$ for all model parameters fell between 0.99 and 1.01, and the effective sample size was greater than 2000 for all parameters. Autocorrelation was low among consecutive thinned observations (*r*<0.10 in all models). We viewed trace plots for fixed and random effects to ensure appropriate sampling of the posterior distribution, and we ran three chains of each model to check for model convergence using the Gelman–Rubin diagnostic (Gelman and Rubin, 1992). We report our model estimates as the mean of the posterior distributions ±95% highest density credible intervals (CI).
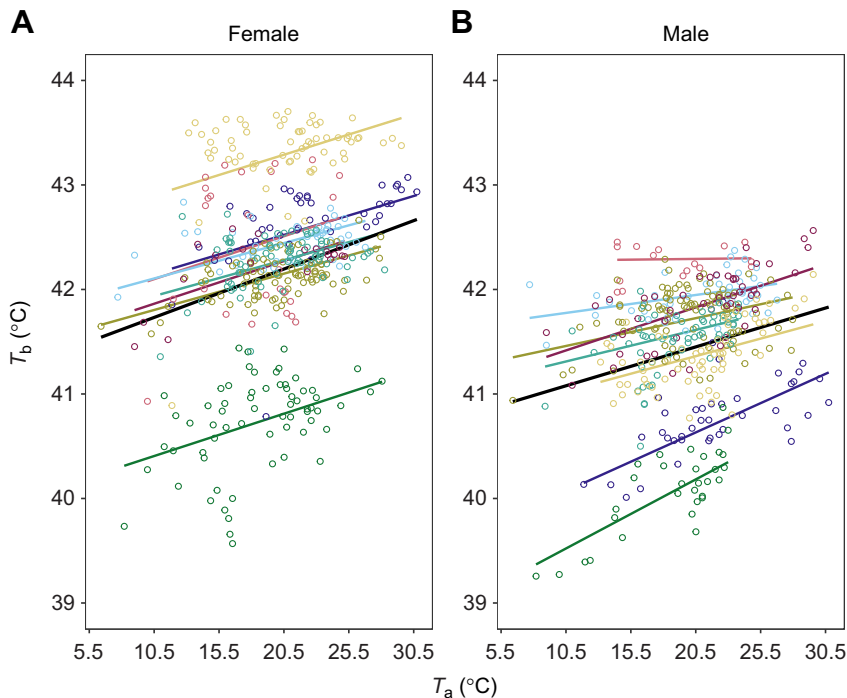
Table 3. Differences between sexes in daytime subcutaneous body temperature (T_b) of control female tree swallows: random effects

	Female			Male		
	Estimate	95% CI [LCI, UCI]	ESS	Estimate	95% CI [LCI, UCI]	ESS
Variance components						
$V_{ind\ intercept}$	0.87	0.39, 1.47	2326	0.82	0.40, 1.38	2646
$V_{ind\ slope}$	0.10	0.00, 0.24	2423	0.28	0.10, 0.56	2391
V_{cov}	0.04	−0.87, 1.00	2624	−0.65	−1.00, 0.01	2465
V_e	0.26	0.25, 0.28	2557	0.81	0.74, 0.90	2409
Repeatability						
$R_{intercept}$	0.89	0.77, 0.98	-	0.92	0.82, 0.98	-
R_{slope}	0.14	0.00, 0.45	-	0.58	0.25, 0.92	-

Variance components and repeatability estimates of T_b according to sex. n_{female}=8, n_{male}=8. Note that repeatability estimates are 'adjusted' (Nakagawa and Schielzeth, 2010) to control for the influence of population-level predictors (e.g. T_a) on mean T_b. $V_{ind,intercept}$, among-individual variance in the intercept; $V_{ind,slope}$, among-individual variance in the slope; V_{cov}, covariance between intercept and slopes; V_e, residual variance.

Fig. 2. T_b of provisioning tree swallows across T_a, as predicted from a Bayesian hierarchical model. (A) Females and (B) males. Note that only control birds (n_{female}=8, n_{male}=8) were included in this analysis. In both groups, there was a shallow increase in T_b with T_a at the population level (black lines). Both groups showed consistent among-individual variation in the intercept of T_b (T_a=0, ~20.5°C), but the degree of among-individual variation in the slope of T_b was much greater in males than in females. Each coloured line represents the predicted T_b of one bird, and dots represent each bird's respective raw data. Pairs are colour coded (i.e. dark green line in each plot represents the female and male of the same nest).

We calculated repeatability estimates using the variance estimates of the intercepts and slopes, according to Nakagawa and Schielzeth (2010). Specifically, $R_{intercept}$ and R_{slope} were calculated as follows:

$$R_{intercept} = \frac{V_{ind\ intercept}}{V_{ind\ intercept} + V_e} \quad (1)$$

$$R_{slope} = \frac{V_{ind\ slope}}{V_{ind\ slope} + V_e} \quad (2)$$

where $V_{ind\ intercept}$ is the among-individual variance in the intercept, $V_{ind\ slope}$ is the among-individual variance in the slope, and V_e is the residual variance among days. Given that temperature was mean-centred and standardized to 2 s.d. units, repeatability of the intercept was calculated at 20.58°C, which is close to the average temperature at our study site for the period of study (May–July 2018, mean±s.d. 18.7±6.3°C). Note that our repeatability estimates at the intercept are conditional, in that they would change based on where they are calculated along the temperature gradient (Nakagawa and Schielzeth, 2010).

RESULTS
Repeatability of T_b by sex and T_a reaction norms
The grand mean T_b averaged across all control female and male swallows (n=16) was 42.1°C at the intercept (T_a=0, or ~21°C), with T_b ranging from 41.07 to 43.8°C (Fig. 1A). On average, the mean T_b of female and male swallows increased with T_a (β=0.38, 95% CI [0.24–0.51]) and feeding rate (β=0.11 [0.07–0.15]), and there was strong evidence that females were ~0.8°C warmer than males (i.e. sex; Table 2; 97% posterior probability distribution<0), as indicated by the strength of the coefficients, associated credible intervals, and the percentage of posterior probability distribution that fell below 0. However, there was no evidence that the sexes differed in mean T_b as a function of T_a, at the population level (i.e. sex×T_a, β=−0.03 [−0.30–0.24]).

Repeatability of average T_b was high, but similar, for the two sexes (i.e. $R_{intercept}$; Table 3), and hypothesis tests confirmed no evidence

for a difference between female and male T_b repeatability (β=−0.07 [−0.18, 0.05], K=0.18). However, there was substantial evidence that females and males differed in their mean slope repeatability (Table 3, $R_{slope,female}$=0.14, $R_{slope,male}$=0.58; Fig. 2), as indicated by the Bayes factor and posterior probability (hypothesis test, β=0.34 [−0.07, 0.68], K=11.8, posterior probability=92%). Additionally, the degree of phenotypic flexibility (i.e. variance in reaction norm slopes) varied between sexes (Levene's test, $F_{1,14}$=7.63, P=0.02). Taken together, these results provide strong evidence that males displayed greater among-individual variation in their T_b across T_a compared with females. Relatedly, there was a negative relationship between intercepts and slopes in males (Fig. 2), but no evidence for a relationship in females (i.e. V_{cov}; Table 3). For males, this means that individuals with lower average T_b tended to be more responsive to changes in T_a compared with individuals with a higher average T_b (i.e. negative covariance).

Repeatability of T_b by treatment and T_a reaction norms
The grand mean T_b averaged across all female swallows (n=16) was 42.1°C at T_a≈21°C, but individuals differed from each other by as much as 2.49°C (Fig. 1B; range 40.80–43.29°C). The mean T_b of female swallows increased with T_a (β=0.36 [0.23–0.47]) and feeding rate (β=0.14 [0.10–0.18]), with some evidence that T_b was lower in trimmed compared with control birds (i.e. treatment, β=−0.54 [−1.43, 0.33]; 88% posterior probability distribution<0) (Table 4). Further, while control and trimmed birds maintained relatively similar T_b across T_a (i.e. treatment×T_a, β=−0.10 [−0.29, 0.11]; Fig. 3), 85% of the posterior probability distribution was <0, suggesting that trimmed birds had, on average, shallower slopes across T_a than control birds.

Repeatability of mean T_b in control and trimmed females was high ($R_{intercept\ control}$=0.89 [0.77, 0.99]; $R_{intercept\ trimmed}$=0.79 [0.59, 0.96]; Table 5), but there was no evidence for differences in repeatability between the treatments (hypothesis test, β=−0.01 [−0.19, 0.20], K=0.88). There was low repeatability in individual slopes in both treatments ($R_{slope\ control}$=0.12 [0.00, 0.41]; $R_{slope\ trimmed}$=0.22 [0.00, 0.55]; Table 5), indicating that individuals did not consistently
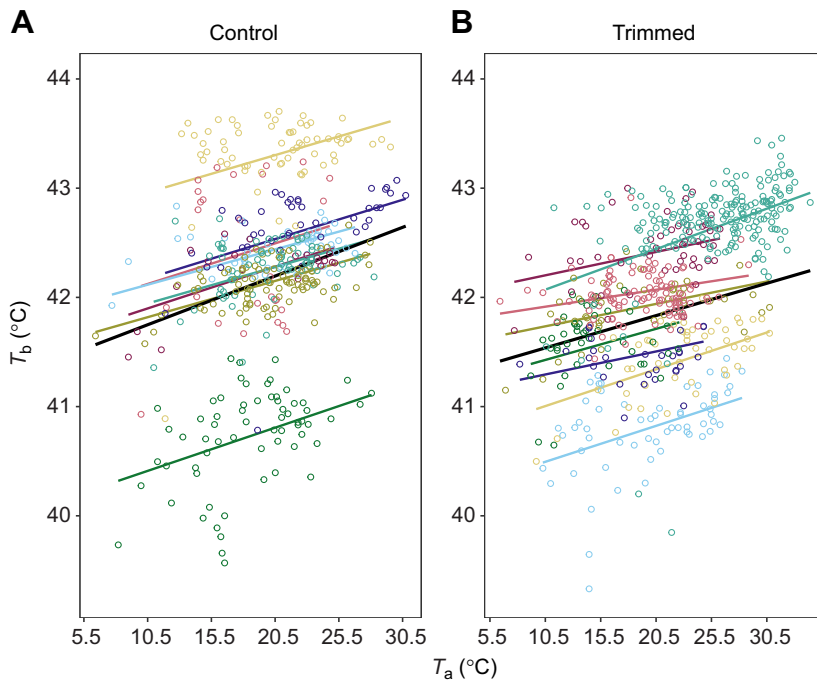
Fig. 3. T_b of female tree swallows without and with an experimental capacity to dissipate body heat. (A) Control birds and (B) trimmed birds ($n_{control}$=8, $n_{trimmed}$=8). In both groups, there was a shallow increase in T_b with T_a at the population level (black lines) and consistent among-individual variation in the intercept of T_b (T_a=0, ~20.5°C). However, there was low repeatability in the slope of the T_b response across T_a. Each coloured line represents the predicted T_b of one bird, and dots represent each bird's respective raw data. Predicted trend lines are derived from a Bayesian hierarchical model.

maintain differences in T_b across T_a, irrespective of treatment. Additionally, we found no evidence for a difference in treatment slope repeatability (hypothesis test, β=−0.20 [−0.56, 0.15], K=0.16). However, the degree of phenotypic flexibility (i.e. variance in reaction norm slope) varied between treatments (Levene's test, $F_{1,14}$=12.38, P=0.003).

We found no support for a relationship between individual identity intercepts and slopes ($V_{cov\ control}$=−0.05 [−0.97, 0.88];

$V_{cov\ trimmed}$=−0.24 [−1.00, 0.59], Table 5), in both treatments, meaning that an individual's average T_b and the plasticity of T_b (across changes in T_a) were not related.

DISCUSSION
Average T_b is repeatable for both sexes, but only males show repeatable reaction norms
We found strong evidence that average T_b is a repeatable trait in female and male tree swallows, meaning that in both sexes, individuals consistently differed from each other in their mean daytime T_b. We also found evidence that T_b was flexible across T_a, and this flexibility was more repeatable in males than in females. The reason for the greater repeatability in the male reaction norm slope is unclear but may be related to sex-specific differences in life-history strategies. While the feeding rate, and variation in feeding rate, was similar between males and females in our study (mean±s.d. 10.4±2.4 females, 10.9±2.7 males), males may be more variable in their overall contribution to parental care than females, which may be reflected in other parental care metrics not measured here (e.g. food load: du Plessis et al., 2012; nest cleaning). Male tree swallows are also known to exhibit polygyny (Dunn and Robertson, 1992), which may affect the time spent attending to their primary nest. Finally, males may be more variable in their body condition,

Table 4. Differences between treatments in daytime T_b of female tree swallows: fixed effects

Term	Estimate	95% CI [LCI, UCI]	ESS
Intercept	42.42	41.71, 43.12	2680
Treatment	−0.54	−1.43, 0.33	2652
T_a	0.36	0.23, 0.47	2618
Lay date	−0.06	−1.59, 1.65	2714
Brood size	0.72	−0.62, 2.17	2751
Feeding rate	0.14	0.10, 0.18	2640
Treatment×T_a	−0.10	−0.29, 0.11	2791

Parameter estimates (±95% CI) for the population-level terms (i.e. fixed effects). $n_{control}$=8, $n_{trimmed}$=8. Note that repeatability estimates are 'adjusted' (Nakagawa and Schielzeth, 2010) to control for the influence of population-level predictors (e.g. T_a) on mean T_b.

Table 5. Differences between treatments in daytime T_b of female tree swallows: random effects

	Control			Trimmed		
	Estimate	95% CI [LCI, UCI]	ESS	Estimate	95% CI [LCI, UCI]	ESS
Variance components						
$V_{ind,intercept}$	0.87	0.38, 1.52	2708	0.64	0.29, 1.11	2511
$V_{ind,slope}$	0.09	0.00, 0.22	2154	0.15	0.01, 0.33	2562
V_{cov}	−0.05	−0.97, 0.88	2657	−0.24	−1.00, 0.59	2757
V_e	0.27	0.25, 0.29	2332	1.08	0.99, 1.19	2358
Repeatability						
$R_{intercept}$	0.89	0.77, 0.99	–	0.79	0.59, 0.96	–
R_{slope}	0.12	0.00, 0.41	–	0.22	0.00, 0.55	–

Variance components and repeatability estimates of T_b according to treatment. $n_{control}$=8, $n_{trimmed}$=8. Note that repeatability estimates are 'adjusted' (Nakagawa and Schielzeth, 2010) to control for the influence of population-level predictors (e.g. T_a) on mean T_b.

or other aspects of morphology that affect insulation (and consequently T_b), such as feather density or length.

We also found that males with lower average T_b tended to have sharper increases in T_b with increasing T_a, compared with males with higher average T_b. One possible explanation is that males with high T_b may be unable (or unwilling) to raise their T_b further because of costs associated heat stress (i.e. heat dissipation limitation hypothesis: Speakman and Król, 2010). This would be consistent with our previous findings showing that limits to sustained energy expenditure in tree swallows may be a consequence of heat dissipation ability at high temperatures (Tapper et al., 2020b).

That we did not see this same effect in females, despite their higher average T_b than males at the population level, may be due to experimental error; for example, if the PIT tag of the female with the highest T_b (Fig. 1A, ID 8F) was implanted in deeper tissue (muscle) than in the other birds, it could raise the population-average T_b in females. While excluding this individual from the analysis reduced the magnitude of the difference between the sexes' average T_b (sex coefficient: 0.85 included, 0.60 excluded), it does not change our conclusions, given that most of the posterior distribution still excluded 0 (96% <0 included, 90% <0 excluded). A biological reason could be that because males lack a brood patch, they are more limited in their ability to dissipate heat relative to females. Therefore, as work rate increases, T_b may rise quicker in males than in females. Alternatively, females and males may differ in their tolerance of thermally induced cellular damage (reviewed in Hansen, 2009; Iossa, 2019; Pérez et al., 2008; Romero-Haro et al., 2016; but see Walsh et al., 2019) and, consequently, females may be less restricted in their ability to raise T_b. This, however, remains speculative in light of research showing no sex-specific differences in oxidative stress in small passerine birds (Costantini et al., 2006, 2007). A final explanation may be that male and female T_b are differentially constrained by other physiological traits. Regardless of the reason, differences in trait correlation between the sexes may imply historical differences in the strength of selection acting on T_b, because selective pressures tend to constrain trait variation within populations across time (Baker et al., 2018; Bell and Sih, 2007; DiRienzo et al., 2016).

Repeatability of female T_b does not differ as a function of relative heat loss

Experimentally manipulating heat loss did not influence the repeatability of average T_b (i.e. reaction norm intercept, 21°C) or T_b across T_a (i.e. reaction norm slope), although it did influence the degree of phenotypic flexibility in T_b, as indicated by different variance in slopes between treatments. That there was greater variance in the T_b slope estimates among trimmed individuals relative to control individuals, but no difference in slope repeatability, suggests that the greater variance in slope steepness observed among trimmed birds is likely due to differences in how individuals respond to environmental factors. For example, sources of variation likely to cause within-individual differences, such as variable environmental conditions (e.g. solar radiation, wind speed, insect abundance), daily fluctuations in internal state (e.g. hormonal levels), short-term behavioural adjustments (e.g. trailing legs during flight, fluffing feathers) and sources of experimental bias (e.g. measurement error, missing data), may have influenced the T_b slope response more in trimmed compared with control birds. Alternatively, it is possible that we did not detect a difference in slope repeatability between treatments because of small sample sizes and a lack of statistical power.

While there were no differences in the repeatability of T_b between treatments, control females were warmer than trimmed females in

average T_b, and their average increase in T_b across T_a (i.e. black line, Fig. 3) was also steeper than in trimmed females. Because of similar repeatability in T_b between treatments, the population-level differences seen in this study are likely to be a consequence of individuals within each group responding consistently to the effect of treatment on T_b. However, it is possible that we may have seen greater inter-individual variation in response to the treatment had we trimmed a larger surface area (i.e. increased the area of heat transfer), because of the difficulty in regulating T_b with increased surface area exposure.

Despite observing low slope repeatability in both groups, there were consistent differences in the way individuals responded to changes in T_a. For instance, some individuals increased their T_b more steeply in response to increasing T_a than others (Fig. 3), suggesting there are different strategies of regulating T_b (Tapper et al., 2020b), or differences in the tolerance of higher T_a (Noakes and McKechnie, 2019). Additionally, the flexibility in the T_b by T_a response could have been influenced by variation in individual energy expenditure, which could arise because of differences in food availability, or the interaction between food availability and T_a (Briga and Verhulst, 2017). Regardless of the reason, such phenotypic flexibility within populations opens the possibility that inter-individual differences in thermoregulatory traits are heritable, and that phenotypic flexibility in T_b may itself be heritable. However, we recognize that even heritability does not necessarily imply trait responsiveness to natural selection (Vatka et al., 2020), and that a variety of phenotypically plastic traits have been reported to be non-responsive to selection by T_a (Arnold et al., 2019).

We note that the repeatability estimates reported here may be not representative of the among-individual variation in swallow core T_b, given that our metric of T_b was peripheral to the core, and consequently is likely to be influenced by both changes in environmental conditions and heat flow across the body, at least to some degree (Nord et al., 2016; Andreasson et al., 2020b). Future studies should consider simultaneously measuring individual variation in both deep and peripheral T_b, remotely, to determine the degree of variability that arises from recording T_b closer to the skin surface. Additionally, it remains to be seen whether among-individual differences in the flexibility of T_b that we report are heritable, and ultimately adaptive.

Data availability
Data are available in Dryad (Tapper, 2021): bcc2fqzbd.

References

Andreasson, F., Hegemann, A., Nord, A. and Nilsson, J.-Å. (2020a). Experimental facilitation of heat loss affects work rate and innate immune function in a breeding passerine bird. *J. Exp. Biol.* **223**, jeb219790. doi:10.1242/jeb.219790

Andreasson, F., Nord, A. and Nilsson, J.-Å. (2020b). Age differences in night-time metabolic rate and body temperature in a small passerine. *J. Comp. Physiol. B* **190**, 349-359. doi:10.1007/s00360-020-01266-5

Angilletta, M. J., Cooper, B. S., Schuler, M. S. and Boyles, J. G. (2010). The evolution of thermal physiology in endotherms. *Front. Biosci.* **2**, 861-881. doi:10.2741/e148

Arnold, P. A., Nicotra, A. B. and Kruuk, L. E. B. (2019). Sparse evidence for selection on phenotypic plasticity in response to temperature. *Philos. Trans. R. Soc B* **374**, 20180185. doi:10.1098/rstb.2018.0185

Auer, S. K., Bassar, R. D., Salin, K. and Metcalfe, N. B. (2016). Repeatability of metabolic rate is lower for animals living under field versus laboratory conditions. *J. Exp. Biol.* **219**, 631-634. doi:10.1242/jeb.133678

Baker, M. R., Goodman, A. C., Santo, J. B. and Wong, R. Y. (2018). Repeatability and reliability of exploratory behavior in proactive and reactive zebrafish, *Danio rerio. Sci. Rep.* **8**, 12114. doi:10.1038/s41598-018-30630-3

Bell, A. M. and Sih, A. (2007). Exposure to predation generates personality in threespined sticklebacks (*Gasterosteus aculeatus*). *Ecol. Lett.* **10**, 828-834. doi:10.1111/j.1461-0248.2007.01081.x

Biro, P. A. and Stamps, J. A. (2010). Do consistent individual differences in metabolic rate promote consistent individual differences in behavior? *Trends Ecol. Evol.* **25**, 653-659. doi:10.1016/j.tree.2010.08.003

Boratyński, J. S., Jefimow, M. and Wojciechowski, M. S. (2017). Individual differences in the phenotypic flexibility of basal metabolic rate in siberian hamsters are consistent on short- and long-term timescales. *Physiol. Biochem. Zool.* **90**, 139-152. doi:10.1086/689870

Boratyński, J. S., Iwińska, K. and Bogdanowicz, W. (2019). An intra-population heterothermy continuum: Notable repeatability of body temperature variation in food-deprived yellow-necked mice. *J. Exp. Biol.* **222**, jeb197152. doi:10.1242/jeb.197152

Boyles, J. G., Thompson, A. B., McKechnie, A. E., Malan, E., Humphries, M. M. and Careau, V. (2013). A global heterothermic continuum in mammals. *Glob. Ecol. Biogeogr.* **22**, 1029-1039. doi:10.1111/geb.12077

Bozinovic, F. (2007). Long-term repeatability of body mass and body temperature (but not basal metabolism) in the free-ranging leaf-eared mouse. *Evo. Ecol. Res.* **9**, 547-554.

Briga, M. and Verhulst, S. (2017). Individual variation in metabolic reaction norms over ambient temperature causes low correlation between basal and standard metabolic rate. *J. Exp. Biol.* **220**, 3280-3289. doi:10.1242/jeb.160069

Broggi, J., Hohtola, E., Koivula, K., Orell, M. and Nilsson, J.-Å. (2009). Long-term repeatability of winter basal metabolic rate and mass in a wild passerine. *Funct. Ecol.* **23**, 768-773. doi:10.1111/j.1365-2435.2009.01561.x

Bürkner, P.-C. (2017). brms: an R Package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1-28. doi:10.18637/jss.v080.i01

Burness, G., Huard, J. R., Malcom, E. and Tattersall, G. J. (2013). Post-hatch heat warms adult beaks: irreversible physiological plasticity in Japenese quail. *Proc. R. Soc. B.* **280**, 20131436. doi:10.1098/rspb.2013.1436

Cai, K. C., van Mil, S., Murray, E., Mallet, J.-F., Matar, C. and Ismail, N. (2016). Age and sex differences in immune response following LPS treatment in mice. *Brain Behav. Immun.* **58**, 327-337. doi:10.1016/j.bbi.2016.08.002

Connolly, M. S. and Lynch, C. B. (1981). Circadian variation of strain differences in body temperature and activity in mice. *Physiol. Behav.* **27**, 1045-1049. doi:10.1016/0031-9384(81)90368-1

Cornelius, E. A., Vézina, F., Regimbald, L., Hallot, F., Petit, M., Love, O. P. and Karasov, W. H. (2017). Chickadees faced with unpredictable food increase fat reserves but certain components of their immune function decline. *Physiol. Biochem. Zool.* **90**, 190-200. doi:10.1086/689913

Costantini, D., Casagrande, S., De Filippis, S., Brambilla, G., Fanfani, A., Tagliavini, J. and Dell'Omo, G. (2006). Correlates of oxidative stress in wild kestrel nestlings (*Falco tinnunculus*). *J. Comp. Physiol. B* **176**, 329-337. doi:10.1007/s00360-005-0055-6

Costantini, D., Cardinale, M. and Carere, C. (2007). Oxidative damage and anti-oxidant capacity in two migratory bird species at a stop-over site. *Comp. Biochem. Phyiol. C* **144**, 363-371. doi:10.1016/j.cbpc.2006.11.005

Dammhahn, M., Landry-Cuerrier, M., Réale, D., Garant, D. and Humphries, M. M. (2017). Individual variation in energy-saving heterothermy affects survival and reproductive success. *Funct. Ecol.* **31**, 866-875. doi:10.1111/1365-2435.12797

Dikmen, S., Cole, J. B., Null, D. J. and Hansen, P. J. (2012). Heritability of rectal temperature and genetic correlations with production and reproduction traits in dairy cattle. *J. Dairy Sci.* **95**, 3401-3405. doi:10.3168/jds.2011-4306

DiRienzo, N., Niemelä, P. T., Hedrick, A. V. and Kortet, R. (2016). Adult bacterial exposure increases behavioral variation and drives higher repeatability in field crickets. *Behav. Ecol. Sociobiol.* **70**, 1941-1947. doi:10.1007/s00265-016-2200-5

du Plessis, K. L., Martin, R. O., Hockey, P. A. R., Cunningham, S. J. and Ridley, A. R. (2012). The costs of keeping cool in a warming world: Implications of high temperatures for foraging, thermoregulation and body condition of an arid-zone bird. *Glob. Chang. Biol.* **18**, 3063-3070. doi:10.1111/j.1365-2486.2012.02778.x

Dunn, P. O. and Robertson, R. J. (1992). Geographic variation in the importance of male parental care and mating systems in tree swallows. *Behav. Ecol.* **3**, 291-299. doi:10.1093/beheco/3.4.291

Gagnon, D. and Kenny, G. P. (2012). Does sex have an independent effect on thermoeffector responses during exercise in the heat? *J. Physiol.* **590**, 5963-5973. doi:10.1113/jphysiol.2012.240739

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statist. Med.* **27**, 2865-2873. doi:10.1002/sim.3107

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7**, 457-472. doi:10.1214/ss/1177011136

Gourdine, J.-L., Mandonnet, N., Giorgi, M. and Renaudeau, D. (2017). Genetic parameters for thermoregulation and production traits in lactating sows reared in tropical climate. *Animal* **11**, 365-374. doi:10.1017/S175173111600135X

Hansen, P. J. (2009). Effects of heat stress on mammalian reproduction. *Phil. Trans. R. Soc. B* **364**, 3341-3350. doi:10.1098/rstb.2009.0131

Hayes, J. P., Bible, C. A. and Boone, J. D. (1998). Repeatability of mammalian physiology: evaporative water loss and oxygen consumption of *Dipodomys merriami. J. Mammal.* **79**, 475-485. doi:10.2307/1382978

Iossa, G. (2019). Sex-specific differences in thermal fertility limits. *Trends Ecol. Evol.* **34**, 490-492. doi:10.1016/j.tree.2019.02.016

Jeffreys, H. (1961). *The Theory of Probability.* Oxford University Press.

Kisliouk, T., Cramer, T. and Meiri, N. (2017). Methyl CpG level at distal part of heat-shock protein promoter HSP70 exhibits epigenetic memory for heat stress by modulating recruitment of POU2F1-associated nucleosome-remodeling deacetylase (NuRD) complex. *J. Neurochem.* **141**, 358-372. doi:10.1111/jnc.14014

Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Adv. Meth. Pract. Psych. Sci.* **1**, 270-280. doi:10.1177/2515245918771304

Lee, M. D. and Wagenmakers, E. J. (2013). *Bayesian Cognitive Modeling: A Practical Course.* Cambridge University Press.

Li, X., Nie, C., Liu, Y., Chen, Y., Lv, X., Wang, L., Zhang, J., Yang, W., Li, K., Zheng, C. et al. (2020). The genetic architecture of early body temperature and its correlation with *Salmonella pullorum* resistance in three chicken breeds. *Front. Genet.* **10**, 1287. doi:10.3389/fgene.2019.01287

Møller, A. P. (2010). Body temperature and fever in a free-living bird. *Comp. Biochem. Physiol. B* **156**, 68-74. doi:10.1016/j.cbpb.2010.02.006

Nakagawa, S., Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biol. Rev.* **85**, 935-956. doi:10.1111/j.1469-185X.2010.00141.x

Nespolo, R. F. and Franco, M. (2007). Whole-animal metabolic rate is a repeatable trait: a meta-analysis. *J. Exp. Biol.* **210**, 2000-2005. doi:10.1242/jeb.02780

Nespolo, R. F., Bacigalupe, L. D. and Bozinovic, F. (2003). Heritability of energetics in a wild mammal, the leaf-eared mouse (*Phyllotis darwini*). *Evolution* **57**, 1679-1688. doi:10.1111/j.0014-3820.2003.tb00373.x

Nespolo, R. F., Verdugo, C., Cortés, P. A. and Bacigalupe, L. D. (2010). Bioenergetics of torpor in the microbiotherid marsupial, monito del monte (*Dromiciops gliroides*): the role of temperature and food availability. *J. Comp. Physiol. B* **180**, 767-773. doi:10.1007/s00360-010-0449-y

Noakes, M. J. and McKechnie, A. E. (2019). Reaction norms for heat tolerance and evaporative cooling capacity do not vary across a climatic gradient in a passerine bird. *Comp. Biochem. Phyiol. A Mol. Integr. Physiol.* **236**, 110522. doi:10.1016/j.cbpa.2019.06.022

Noakes, M. J. and McKechnie, A. E. (2020). Phenotypic flexibility of metabolic rate and evaporative water loss does not vary across a climatic gradient in an Afrotropical passerine bird. *J. Exp. Biol.* **223**, jeb220137. doi:10.1242/jeb.220137

Nord, A. and Giroud, S. (2020). Lifelong effects of thermal challenges during development in birds and mammals. *Front. Physiol.* **11**, 419. doi:10.3389/fphys.2020.00419

Nord, A. and Nilsson, J.-Å. (2019). Heat dissipation rate constrains reproductive investment in a wild bird. *Funct. Ecol.* **33**, 250-259. doi:10.1111/1365-2435.13243

Nord, A., Chiriac, S., Hasselquist, D. and Nilsson, J.-Å. (2013). Endotoxin injection attenuates rest-phase hypothermia in wintering great tits through the onset of fever. *Funct. Ecol.* **27**, 236-244. doi:10.1111/1365-2435.12003

Nord, A., Lehmann, M., MacLeod, R., McCafferty, D. J., Nager, R. G., Nilsson, J.-Å. and Helm, B. (2016). Evaluation of two methods for minimally invasive peripheral body temperature measurement in birds. *J. Avian Biol.* **47**, 417-427. doi:10.1111/jav.00845

O'Connor, R. S., Le Pogam, A., Young, K. G., Robitaille, F., Choy, E. S., Love, O. P., Elliott, K. H., Hargreaves, A. L., Berteaux, D., Tam, A. et al. (2021). Limited heat tolerance in an arctic passerine: thermoregulatory implications for cold-specialized birds in a rapidly warming world. *Ecol. Evol.* **11**, 1609-1619. doi:10.1002/ece3.7141

Oswald, K. N., Evlambiou, A. A., Ribeiro, Â. M. and Smit, B. (2018). Tag location and risk assessment for passive integrated transponder-tagging passerines. *Ibis* **160**, 453-457. doi:10.1111/ibi.12558

Pessato, A., McKechnie, A. E., Buchanan, K. L. and Mariette, M. M. (2020). Vocal panting: A novel thermoregulatory mechanism for enhancing heat tolerance in a desert-adapted bird. *Sci. Rep.* **10**, 18914. doi:10.1038/s41598-020-75909-6

Pérez, J. H., Ardia, D. R., Chad, E. K. and Clotfelter, E. D. (2008). Experimental heating reveals nest temperature affects nestling condition in tree swallows (*Tachycineta bicolor*). *Biol. Lett.* **4**, 468-471. doi:10.1098/rsbl.2008.0266

Romero-Haro, A. A., Sorci, G. and Alonso-Alvarez, C. (2016). The oxidative cost of reproduction depends on early development oxidative stress and sex in a bird species. *Proc. R. Soc. B* **283**, 1833. doi:10.1098/rspb.2016.0842

Rønning, B., Moe, B. and Bech, C. (2005). Long-term repeatability makes basal metabolic rate a likely heritable trait in the zebra finch *Taeniopygia guttata*. *J. Exp. Biol.* **208**, 4663-4669. doi:10.1242/jeb.01941

Rønning, B., Jensen, H., Moe, B. and Bech, C. (2007). Basal metabolic rate: Heritability and genetic correlations with morphological traits in the zebra finch. *J. Evol. Biol.* **20**, 1815-1822. doi:10.1111/j.1420-9101.2007.01384.x

Ruuskanen, S., Hsu, B. Y. and Nord, A. (2021). Endocrinology of thermoregulation in birds in a changing climate. *Mol. Cell. Endocrinol.* **519**, 111088. doi:10.1016/j.mce.2020.111088

Schoenemann, K. L. and Bonier, F. (2018). Repeatability of glucocorticoid hormones in vertebrates: a meta-analysis. *PeerJ* **6**, e4398. doi:10.7717/peerj.4398

Schreier, K. C. and Grindstaff, J. L. (2020). Repeatable behavioural and immune defence strategies against infection are not traded off. *Anim. Behav.* **162**, 11-22. doi:10.1016/j.anbehav.2020.01.008

Smit, B., Harding, C. T., Hockey, P. A. R. and McKechnie, A. E. (2013). Adaptive thermoregulation during summer in two populations of an arid-zone passerine. *Ecology* **94**, 1142-1154. doi:10.1890/12-1511.1

Speakman, J. R. and Król, E. (2010). Maximal heat dissipation capacity and hyperthermia risk: neglected key factors in the ecology of endotherms: Heat dissipation limit theory. *J. Anim. Ecol.* **79**, 726-746. doi:10.1111/j.1365-2656.2010.01689.x

Stager, M., Senner, N. R., Tobalske, B. W. and Cheviron, Z. A. (2020). Body temperature maintenance acclimates in a winter-tenacious songbird. *J. Exp. Biol.* **223**, jeb221853. doi:10.1242/jeb.221853

Tapper, S. (2021). Body temperature is a repeatable trait in a free-ranging passerine bird. *Dryad Dataset*, doi:10.5061/dryad.bcc2fqzbd

Tapper, S., Nocera, J. J. and Burness, G. (2020a). Heat dissipation capacity influences reproductive performance in an aerial insectivore. *J. Exp. Biol.* **223**, jeb222232. doi:10.1242/jeb.222232

Tapper, S., Nocera, J. J. and Burness, G. (2020b). Experimental evidence that hyperthermia limits offspring provisioning in a temperate-breeding bird. *R. Soc. Open Sci.* **7**, 201589. doi:10.1098/rsos.201589

Tieleman, B. I., Croese, E., Helm, B. and Versteegh, M. A. (2010). Repeatability and individual correlates of microbicidal capacity of bird blood. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **156**, 537-540. doi:10.1016/j.cbpa.2010.04.011

Van Goor, A., Bolek, K. J., Ashwell, C. M., Persia, M. E., Rothschild, M. F., Schmidt, C. J. and Lamont, S. J. (2015). Identification of quantitative trait loci for body temperature, body weight, breast yield, and digestibility in an advanced intercross line of chickens under heat stress. *Genet. Sel. Evol.* **47**, 96. doi:10.1186/s12711-015-0176-7

Vatka, E., Orell, M., Rytkönen, S. and Merilä, J. (2020). Effects of ambient temperatures on evolutionary potential of reproductive timing in boreal passerines. *J. Anim. Ecol.* **90**, 367-375. doi:10.1111/1365-2656.13370

Walsh, B. S., Parratt, S. R., Atkinson, D., Snook, R. R., Bretman, A. and Price, T. A. R. (2019). Integrated approaches to studying male and female thermal fertility limits. *Trends Ecol. Evol.* **34**, 492-493. doi:10.1016/j.tree.2019.03.005

White, C. R., Schimpf, N. G. and Cassey, P. (2013). The repeatability of metabolic rate declines with time. *J. Exp. Biol.* **216**, 1763-1765. doi:10.1242/jeb.076562