

Supplementary Material

Methods

Animal capture, housing and body temperature measurements

After capture using mist-nets, the birds were temporarily held in six-person tents (Hobitat 6, REI, Inc., USA), 3.0 x 2.5 x 1.8 m, adapted as aviaries set up at a field laboratory. Tents were equipped with string and natural perches, and an *ad libitum* supply of artificial nectar (Nektar Plus, Nekton, Germany) in 50-mL plastic syringes during the birds' time in captivity. Food was withheld from 30 min before dark, at which time birds were transferred into individual roosting enclosures, each consisting of a ~20-cm diameter nylon mesh sleeve outfitted with a wire perch.

After being weighed to 0.01 g using a Jennings Jscale Mack 20 electronic balance, a 36-gauge thermocouple probe (RET-4, Physitemp, Clifton NJ, USA) was inserted 1-2 cm into the cloaca of each bird, at which depth where a slight withdrawal did not result in a change in the T_b reading. Each thermocouple was held in place using a small piece of lab tape fastened to the of the bird's central retrices. The thermocouples were connected to a 12-channel thermocouple reader/data logger (Omega RDXL12, Omega Engineering, Norwalk CT, USA), as were additional thermocouples used to record air temperature (T_a) at the same height at the bird. Prior to the study, the reader/data logger and thermocouples were calibrated against a thermometer with NIST-traceable accuracy. Data for T_b and T_a were recorded at 10-s intervals and logged prior to download and analysis the following morning after birds had rewarmed. Birds weighed again within 30 minutes of rewarming.

Statistical analyses

We modelled the response of two torpor variables, minimum T_b and overnight M_b loss. To do this, we first confirmed normality and homoscedasticity of the data, then fit generalized linear multilevel models using the *brms* [1, 2] and *stan* [3] packages in R [4]. We considered overnight mass loss to be an index of energy savings associated with torpor, following numerous other authors [e.g., [5]. Estimating energy expenditure from mass loss is complicated by hummingbirds' proportionately large crops which may contain nectar of varying concentrations [6], nocturnal metabolism being fuelled by both fat and carbohydrates, and changes in M_b associated with factors other than changes in fat content [7]. Indeed, our M_b loss data reiterate the latter point, as assuming that these losses represented fat alone would yield unrealistic estimates of overnight average metabolic rates, 10 X higher than allometrically predicted basal values [8].

For minimum T_b and overnight M_b loss, we compared four models: (1) bout duration as a fixed effect with no random effects; (2) bout duration as a fixed effect with a species random effect; (3) bout duration as a fixed effect with a phylogenetic random effect; and (4) bout duration as a fixed effect with both species and phylogenetic random effects. We fit models using the *brms()* function with the 'Gaussian' family and default priors. For each model, we ran four chains for 35,000 iterations per chain. After a burn-in of 10,000 iterations, we sampled chains every 10 iterations for a total of 10,000 post burn-in samples per model. We assessed convergence through diagnostics and visual examination of traceplots. For each response, we compared models using the leave-one-out cross-validation information criteria (LOOIC) in the R package *loo* [9].

We visualized minimum T_b and overnight M_b loss across the phylogeny of the species in our study using a published phylogeny of hummingbirds [10] and the *contMap()* function in the R package *phytools* [11]. Furthermore, we estimated the phylogenetic signal in minimum T_b and overnight M_b loss in three ways. First, we estimated Pagel's λ using Bayesian inference in the R package *motmot* [12]. We ran a single chain for 20,000 iterations,

sampling every 10 iterations after a burn-in of 10,000 iterations. Second, we estimated Bloomberg's K using maximum likelihood using the `phylosig()` function in *phytools* [11]. Finally, we followed the approach of Barrow et al. [13] and Lynch [14] and quantified phylogenetic signal from our generalized linear multilevel models that included a phylogenetic random effect by estimating the proportion of total variance attributed to phylogeny. Similarly, we estimated the proportion of total variance attributed to species from models including a species random effect.

References

- [1] Bürkner, P.-C. 2017 brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* **80**, 1-28.
- [2] Bürkner, P.-C. 2018 Advanced Bayesian multilevel modeling with the R package brms. *The R Journal* **10**, 395-411.
- [3] Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. & Riddell, A. 2017 Stan: A probabilistic programming language. *Journal of Statistical Software* **76**.
- [4] Team, R.C. 2018 *R: A language and environment for statistical computing*. Vienna, R Foundation for Statistical Computing.
- [5] Bech, C., Abe, A.S., Steffensen, J.F., Berger, M. & Bicudo, J.E.P.W. 1997 Torpor in three species of Brazilian hummingbirds under semi-natural conditions. *Condor* **99**, 780-788.
- [6] Hainsworth, F.R. & Wolf, L.L. 1972 Crop volume, nectar concentration and hummingbird energetics. *Comp Biochem Physiol*, 359-366.
- [7] Powers, D.R. 1991 Diurnal variation in mass, metabolic rate, and respiratory quotient in Anna's and Costa's hummingbirds. *Physiol Zool* **64**, 850-870.
- [8] Londoño, G.A., Chappell, M.A., del Rosario Castañeda, M., Jankowski, J.E. & Robinson, S.K. 2015 Basal metabolism in tropical birds: latitude, altitude, and the "pace of life". *Functional Ecology* **29**, 338-346.
- [9] Vehtari, A., Gelman, A. & Gabry, J. 2017 Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing* **27**, 1413-1432.
- [10] McGuire, J.A., Witt, C.C., Remsen Jr, J., Corl, A., Rabosky, D.L., Altshuler, D.L. & Dudley, R. 2014 Molecular phylogenetics and the diversification of hummingbirds. *Curr. Biol.* **24**, 910-916.
- [11] Revell, L.J. 2012 phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**, 217-223.
- [12] Puttick, M.N., Ingram, T., Clarke, M. & Thomas, G.H. 2020 MOTMOT: Models of trait macroevolution on trees (an update). *Methods in Ecology and Evolution* **11**, 464-471.
- [13] Barrow, L.N., McNew, S.M., Mitchell, N., Galen, S.C., Lutz, H.L., Skeen, H., Valqui, T., Weckstein, J.D. & Witt, C.C. 2019 Deeply conserved susceptibility in a multi-host, multi-parasite system. *Ecology letters* **22**, 987-998.
- [14] Lynch, M. 1991 Methods for the analysis of comparative data in evolutionary biology. *Evolution* **45**, 1065-1080.

Supplementary Table 1. Comparison of generalized linear multilevel models of minimum body temperature (T_b). Models varied in whether they included bout duration, minimum ambient temperature (T_a), or both as fixed effects and incorporated species, phylogenetic, both or no random effects. The estimated effect size and 95% highest posterior density (HPD) are listed for each effect. The proportion of variance explained and 95% HPD are indicated for models with species and phylogenetic random effects. Model fit was assessed using leave-one-out cross-validation (LOOIC). The difference between each model and the best-fit model is shown as Δlpld (expected log predictive density) with standard error (se). The structure of the full model is as follows: $\text{Min } T_b \sim \text{bout duration} + \text{min } T_a + \text{species} + \text{phylogeny}$.

Response	Fixed effects		Random effects		% variance explained		Δlpld (se)
	Bout duration	Min T_a	Species	Phylogeny	Species	Phylogeny	
Min. T_b	-0.62 (-0.81, -0.43)	0.31 (-0.09, 0.72)	1.68 (0.64, 4.06)	–	0.62 (0.22, 0.93)	–	0
Min. T_b	-0.60 (-0.79, -0.40)	0.34 (-0.07, 0.74)	–	0.42 (0.14, 1.08)	–	0.14 (0.01, 0.49)	-0 (0.4)
Min. T_b	-0.61 (-0.81, -0.41)	0.31 (-0.09, 0.72)	1.35 (0.06, 4.15)	0.36 (0.01, 1.16)	0.47 (0.003, 0.94)	0.12 (0.0002, 0.54)	-0.1 (0.3)
Min. T_b	-0.60 (-0.80, -0.40)	–	1.54 (0.08, 4.75)	0.39 (0.02, 1.29)	0.51 (0.004, 0.95)	0.12 (0.0002, 0.57)	-0.3 (1.7)
Min. T_b	-0.60 (-0.80, -0.41)	–	1.78 (0.65, 4.24)	–	0.63 (0.21, 0.94)	–	-0.4 (1.7)
Min. T_b	-0.58 (-0.78, -0.38)	–	–	0.47 (0.15, 1.24)	–	0.15 (0.01, 0.57)	-0.4 (1.7)
Min. T_b	-0.68 (-0.86, -0.49)	0.31 (-0.18, 0.80)	–	–	–	–	-7.3 (3.5)
Min. T_b	-0.66 (-0.85, -0.47)	–	–	–	–	–	-7.5 (3.5)
Min. T_b	–	0.20 (-0.40, 0.82)	–	0.59 (0.24, 1.41)	–	0.11 (0.02, 0.40)	-13.8 (4.6)
Min. T_b	–	0.19 (-0.42, 0.82)	1.72 (0.07, 5.55)	0.53 (0.04, 1.61)	0.39 (0.001, 0.91)	0.10 (0.0004, 0.46)	-14.3 (4.5)
Min. T_b	–	0.18 (-0.45, 0.81)	2.48 (1.03, 5.55)	–	0.60 (0.22, 0.91)	–	-14.3 (4.5)
Min. T_b	–	0.10 (-0.72, 0.92)	–	–	–	–	-24.0 (4.8)

Supplementary Table 2. Comparison of generalized linear multilevel models of percent body mass (M_b) loss. Models varied in whether they included bout duration, minimum body temperature (T_b), or both as fixed effects and incorporated species, phylogenetic, both or no random effects. The estimated effect size and 95% highest posterior density (HPD) are listed for each effect. The proportion of variance explained and 95% HPD are indicated for models with species and phylogenetic random effects. Model fit was assessed using leave-one-out cross-validation (LOOIC). The difference between each model and the best-fit model is shown as Δelpd (expected log predictive density) with standard error (se). The structure of the full model is as follows: $\% M_b \text{ loss} \sim \text{bout duration} + \text{min } T_b + \text{species} + \text{phylogeny}$.

Response	Fixed effects		Random effects		% variance explained		Δelpd (se)
	Bout duration	T_b min	Species	Phylogeny	Species	Phylogeny	
M_b loss (%)	-0.36 (-0.87, 0.14)	0.43 (-0.19, 1.03)	–	–	–	–	0
M_b loss (%)	-0.64 (-0.96, -0.32)	–	–	–	–	–	-0 (1.4)
M_b loss (%)	–	0.77 (0.37, 1.16)	–	–	–	–	-0.1 (1.0)
M_b loss (%)	-0.34 (-0.89, 0.20)	0.40 (-0.31, 1.09)	–	0.29 (0.01, 0.97)	–	0.02 (0.00002, 0.13)	-1.1 (1.6)
M_b loss (%)	-0.57 (-0.95, -0.19)	–	–	0.32 (0.01, 1.05)	–	0.03 (0.00002, 0.15)	-1.2 (2.3)
M_b loss (%)	-0.60 (-0.97, -0.23)	–	1.18 (0.04, 3.85)	–	0.19 (0.0002, 0.70)	–	-1.2 (2.0)
M_b loss (%)	–	0.72 (0.22, 1.18)	–	0.31 (0.01, 1.07)	–	0.02 (0.00002, 0.15)	-1.3 (1.8)
M_b loss (%)	-0.35 (-0.92, 0.21)	0.40 (-0.32, 1.10)	1.07 (0.03, 3.56)	–	0.16 (0.0002, 0.67)	–	-1.4 (1.6)
M_b loss (%)	–	0.73 (0.24, 1.20)	1.15 (0.04, 3.93)	–	0.18 (0.0003, 0.70)	–	-1.4 (1.6)
M_b loss (%)	-0.56 (-0.96, -0.14)	–	1.31 (0.05, 4.60)	0.36 (0.01, 1.27)	0.21 (0.0003, 0.78)	0.03 (0.00002, 0.20)	-1.6 (2.7)
M_b loss (%)	–	0.69 (0.16, 1.20)	1.32 (0.05, 4.80)	0.36 (0.01, 1.29)	0.21 (0.0003, 0.79)	0.03 (0.00002, 0.21)	-1.7 (2.2)
M_b loss (%)	-0.34 (-0.94, 0.27)	0.37 (-0.41, 1.14)	1.25 (0.01, 1.28)	0.35 (0.01, 1.28)	0.20 (0.0002, 0.76)	0.03 (0.00002, 0.21)	-2.0 (2.3)