

# Individual variation in heat substitution: is activity in the cold energetically cheaper for some individuals than others?

Caroline M Maloney and Vincent Careau\*

Department of Biology, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada

\*Author for Correspondence: [vcareau@uottawa.ca](mailto:vcareau@uottawa.ca)

**KEYWORDS:** activity-thermoregulatory heat substitution, thermoregulation, energy compensation, repeatability

## Summary

We found consistent individual differences in heat substitution, an important yet overlooked mechanism that allows endotherms that are active in the cold to reduce the total energetic cost of activity and thermoregulation.

## ABSTRACT

In many endotherms, a potentially important yet often overlooked mechanism to save energy is the use of the heat generated by active skeletal muscles to replace heat that would have been generated by thermogenesis (i.e., “activity-thermoregulatory heat substitution”). While substitution has been documented numerous times, the extent of individual variation in substitution has never been quantified. Here, we used a home-cage respirometry system to repeatedly measure substitution through the concomitant monitoring of metabolic rate (MR) and locomotor activity in 46 female white-footed mice (*Peromyscus leucopus*). A total of 117 measures of substitution were taken by quantifying the difference in the slope of the relationship between MR and locomotor activity speed at two different ambient temperatures. Consistency

repeatability ( $\pm$ se) of substitution was  $0.313 \pm 0.131$  – hence, about a third of the variation in substitution occurs at the among-individual level. Body length and heart mass were positively correlated with substitution whereas surface area was negatively correlated with substitution. These two sub-organismal traits accounted for the majority of the among-individual variation (i.e., individual differences in substitution were not significant after accounting for these traits). Overall, our results imply that the energetic cost of activity below the thermoneutral zone is consistently cheaper from some individuals than others, and that the energy saved from substitution might be available to invest in fitness-enhancing activities.

## INTRODUCTION

Cold environments represent a unique challenge for endothermic animals. Endotherms can only maintain their high core body temperature ( $T_b$ ) without any homeostatic responses within a small range of ambient temperatures (Hedrick and Hillman, 2016). When the ambient temperature ( $T_a$ ) falls below an endotherm's thermoneutral zone (TNZ), it must engage in thermogenesis to maintain its  $T_b$  constant. This represents an additional energetic cost, and to avoid energy shortfalls endotherms can either increase their energy intake (Arnold et al., 2006) or reduce the cost of thermogenesis (Hetem et al., 2016; Levesque et al., 2016). The energetic cost of thermoregulation can be reduced through various physiological and behavioural mechanisms. Many birds and mammals save energy by using torpor—a state of decreased physiological activity accompanied by lower core  $T_b$ —either in a daily or seasonal pattern (also known as hibernation) (Geiser, 2020). Endotherms can also avoid cold exposure, such as red squirrels in boreal forests who stay in their well-insulated nests instead of foraging on the coldest days (Menzies et al., 2020).

There are situations in which it is impossible for endotherms to use torpor and cold avoidance behavioural mechanisms (Maresh et al., 2015). In fact, it has been estimated that most temperate endotherms routinely experience  $T_a$  that are lower than their TNZ (Humphries and Careau, 2011). In these circumstances, a potentially important yet understudied energy-saving mechanism is the activity-thermoregulatory heat substitution (hereafter, substitution). Substitution occurs when heat generated from physical activity (work) replaces the heat that a resting endotherm below the TNZ would have had to produce through thermogenesis to maintain

its  $T_b$  (Lefèvre and Auget, 1931). The mechanism by which substitution occurs is thought to be the diversion of heat from active skeletal muscle to the body's core (Liwanag et al., 2009). The central benefit of substitution is to reduce overall energetic costs in the cold for an endotherm. In fact, substitution can outright eliminate the cost of activity (COA) (Humphries and Careau, 2011). Thermogenesis represents a sunk cost for an endotherm in the cold, because it must expend energy to compensate for the difference between  $T_a$  and  $T_b$  regardless of the form of thermogenesis (shivering or non-shivering). Substitution cannot reduce the energy used to thermoregulate, but can use that energy efficiently if the muscle activity that generates heat can, in turn, be potentially invested in physiological processes that improve fitness. For example, locomotor activity is integral to resource acquisition through foraging effort, and mating success involves many processes (e.g., courtship, fighting and chasing) that elevate metabolic rate (MR) and require skeletal muscle involvement. By making activity energetically cheaper, substitution may ease energy constraints on an organism, allowing for higher energy allocation towards fecundity, somatic growth, or survival (Stearns, 2000).

Quantifying substitution requires extensive time and effort because it involves measuring MR in resting vs active animals at different temperatures. Perhaps the most straightforward method of estimating substitution is to compare the COA between two ambient temperatures: a warm  $T_a$  ( $COA_w$ ) and a relatively cold  $T_a$  ( $COA_c$ ). If no substitution occurs,  $COA_w$  and  $COA_c$  will be the same (Fig. 1, solid lines). In other words, the slopes of the relationship between MR and activity are parallel at these two measured temperatures, and MR increases by the same amount for each unit activity. If substitution occurs at the lower  $T_a$ , however, some (or all) of the additional energy required for thermoregulation is “paid for” by the heat produced by activity, such that MR does not increase by the same amount for each unit activity (i.e., the slope becomes shallower and  $COA_c$  is lower; Fig. 1, dotted line). Therefore, substitution can be quantified as the difference between  $COA_w$  and  $COA_c$  (Fig. 1). Note that it is possible that the COA is higher at lower  $T_a$  (e.g., due to the disruption of the boundary layer; Pauls, 1981), in which case calculating substitution as  $COA_w - COA_c$  would return a negative value.

Research on substitution has been primarily focused on quantifying the extent of substitution at the population level and how substitution is affected or driven by the mode and intensity of activity and traits such as thermal conductance. Substitution is reduced or absent in

endotherms when they use an atypical mode of locomotion (e.g., terrestrial locomotion in birds) (Bryant et al., 1985; Pohl and West, 1973) and have an increased thermal conductance when active than when resting due to the state of the boundary layer (Hart, 1950; Hart and Heroux, 1955; Pauls, 1981). Morphology and behaviour both play important roles in substitution; for example, sea lions use their extensive peripheral fat stores to reduce conductance and avoid high swim speeds that would likewise increase conductance (Liwanag et al., 2009). Moreover, substitution seems more likely to occur under voluntary activity (Chappell and Hammond, 2004; Chappell et al., 2004) than forced activity (Hart and Heroux, 1955; Pauls, 1981; Yousef et al., 1973). Humphries and Careau (2011) conducted a meta-analysis of substitution studies and found that substitution is negatively correlated with size in birds (but not in mammals) and positively correlated with intensity of activity in both birds and mammals. Overall, substitution is a common, potentially important avenue of saving energy subject to extensive interspecific variation, yet we still do not know the extent and consistency of individual differences in substitution.

Individual variation plays a key role in the adaptive evolution of any trait, as it represents the “raw material” upon which natural selection acts. Moreover, heritable individual differences are a prerequisite for the adaptive divergence among populations, species, and taxa (Hayes and Jenkins, 1997). Given the potential importance of substitution for fitness in endotherms, it is surprising that no study has attempted to quantify whether individuals consistently differ in the degree to which they substitute heat. In other words, we still do not know if substitution is a potentially heritable trait and therefore whether selection (directly or indirectly) could act on this trait. If substitution is repeatable, then this would imply that the energetic cost of activity below TNZ is consistently cheaper from some individuals than others, which might have impact on fitness. Focussing on individual variation not only represents a first step to understanding the adaptive nature of substitution, but may also help identifying its underlying mechanisms and functional relationships with other traits.

For most ecologically and evolutionary relevant phenotypic traits (e.g., behaviour or physiological parameters), an individual’s expression of the trait is labile and can change in response to environmental conditions, across time, or in response to changes in other traits (Araya-Ajoy et al., 2015). Therefore, singular measurements might not solely represent variation

among individuals. Total phenotypic variance in a trait is the sum of among-individual variance (i.e., the difference in individuals mean trait values due to genetic and permanent environmental factors) and within-individual variance (i.e., the differences in trait value for a single individual due to measurement error, labile environmental effects and acclimation). Individual repeatability, calculated as the ratio of among-individual variance over phenotypic variance, is a useful metric because it gives an idea of how much consistent individual differences compare to variation that occurs within individuals. Repeatability also represents the upper limit to heritability (Boake, 1989; Wolak et al., 2012). When estimating repeatability (and heritability), care must be taken when choosing the appropriate set of covariates; while it might be useful to remove “nuisance” sources of variance (e.g., block or chamber effects), removing variation caused by other, co-varying biological traits remains potentially problematic (Wilson, 2008; Wilson, 2018). In fact, it might be useful to compare repeatability estimates with vs without conditioning on various traits to get an idea of how much of the inter-individual differences in a trait is due to among-individual variation in other, co-varying traits (Roche et al., 2016; Santos et al., 2015).

Among all possible sub-organismal traits underlying variation in substitution, the most likely are those that influence heat dissipation. Insulation (pelage and fat) modulates thermal conductance, so variation in traits like fur density and skin (fat and tissue) density affect how much heat is lost from working muscles. Dry heat transfer through conduction and convection (the primary heat-loss mechanism used by rodents) occurs through a gradient between the surface and the environment. Therefore, the overall body shape of an animal that determines its surface/volume ratio might influence substitution (Mitchell et al., 2017b; Mitchell et al., 2018). Similarly, the size of morphological features— primarily tails in rodents—can enhance dry heat transfer (Škop et al., 2020). Another set of possible covariates to substitution include traits that are energetically expensive by nature. Vital organs that are “functionally significant” such as the heart, liver, and kidneys incur a large ongoing metabolic cost (Konarzewski and Diamond, 1995), and their size in individuals is directly tied to energy availability (Mitchell et al., 2017a). As allocating energy to somatic growth and maintenance can be costly, selection for energy saving mechanisms may occur alongside selection for organ size. Hence, individual variation in these vital organs may covary with individual variation in substitution (although they may also vary according to other factors). Skeletal muscle is also energetically expensive—although muscle tissue has a low metabolic rate per gram at rest, it constitutes the majority of lean body

mass and is highly metabolically demanding when the animal is active (Raichlen et al., 2010). The heat used in substitution also is produced by skeletal muscles (González-Alonso, 2012), which could mean that the size of specific skeletal muscles may covary with substitution both because of potential energy savings and because of the role of heat production.

The main objective of this study was to test whether activity-thermoregulatory heat substitution is repeatable in white-footed mouse (*Peromyscus leucopus*), a small North American rodent. *Peromyscus* mice have been used in previous studies of substitution (Chappell and Hammond, 2004; Chappell et al., 2004) and are appropriate in that they occupy a large thermal habitat and are active year-round (Borniger and Nelson, 2017). To quantify substitution as the  $COA_W - COA_C$ , we measured MR and locomotor activity intensity at two different ambient temperatures: one warm temperature of 22°C that corresponded to the housing conditions of the animals to quantify  $COA_W$ , and a colder temperature of 10°C to quantify  $COA_C$ . To estimate repeatability, we took repeated measures of substitution and used mixed models to partition the phenotypic variance at the among- and within-individual levels. A second objective was to test whether substitution is covarying with a suite of sub-organismal traits such as fur density, surface area, organ size, and hematocrit. Calculating adjusted repeatability after having accounted for these covariates, and comparing it to the consistency repeatability, will provide an idea of how much the sub-organismal traits explain inter-individual differences in substitution.

## METHODS

### Ethics

All procedures were approved by the Animal Care Committee at the University of Ottawa (Protocol BLe-3227-A1) and were in accordance with the Canadian Council on Animal Care's guidelines.

### Study animals

Individually marked (ear punches) adult female white-footed mice, originally purchased from the *Peromyscus* Genetic Stock Center (Columbia, SC USA), were used for the experiment. Mice

were 10-13 months old at the start of the study. Mice were housed in groups of four in standard rat cages (42×21×20 cm) and rodent chow and water were provided *ad libitum* in addition to nesting material and enrichment. The light cycle in the room was maintained with 11 hours of light followed by 13 hours of dark, with a 30-min gradual switches starting at 7:00 (dark to light) and at 17:30 (light to dark). Room temperature was controlled at 22°C as per the standard of the facility.

## Experimental design

On September 24<sup>th</sup> 2020, the experiment started with the objective to have 3 repeated measures of substitution in 48 individuals over 4 months. Since we could only measure 8 individuals at a time (see below), mice were assigned to 6 rotating groups of 8 mice each. Each week a different group was measured while the other mice remained in their home cage. Therefore, there was a period of 5-8 weeks in between repeated tests on a given individual, and the experiment lasted until January 9<sup>th</sup>, 2021. Natural and accidental deaths reduced the number of mice in the study, but all tests where the animal was not removed early or died during either temperature condition were included for analysis (see Table 1 for sample sizes of the different variables).

To measure substitution, mice were individually placed in 1 of 8 metabolic cages for a period of 96 hours (4 days), during which MR and locomotor activity were monitored. Mice were randomly assigned to a different metabolic cage for each test. Metabolic cages were housed within an environmental cabinet set at either 22°C (the temperature the animals were acclimatized to) or 10°C for the first and second 48 h of the test (order of temperature randomized). Mice were not acclimated to the metabolic cages in the environmental cabinet prior to the tests, nor were any observations removed to account for acclimation. Given the order in which ambient temperatures were applied was random, however, any change in locomotor activity caused by the novelty of the metabolic chambers would make our repeatability estimate conservative. A thermometer placed on the cabinet wall adjacent to the second lowest shelf confirmed that the temperature stayed within  $\pm 0.5^{\circ}\text{C}$  of the set temperature. Temperature changes within tests (from 10 to 22 or 22 to 10°C) took approximately 30 min and no data was collected during the equilibration period. The cabinet's light source was set to cycle to 12 hours of light followed by 12 hours of dark. Mice were inspected daily through the glass door of the

cabinet but remained undisturbed while tested. Mice were given rodent chow and water *ad libitum* while in the metabolic cages. No nesting material was provided in these cages, and enrichment was limited to the running wheel and a small shelter space. This protocol resulted in one 48-hour set of metabolic and behavioural observations for each temperature condition for each test, which allowed for calculation of substitution depending on the level of voluntary activity expressed by the animal (see below). Metabolic cages were cleaned every two weeks, between tests. This was repeated until all groups were tested three times.

At the end of each test week, the tested group of mice had blood drawn for hematocrit measures. Blood was taken from the saphenous vein, filling individual capillary tubes. Tubes were placed into a hematocrit centrifuge and spun at a speed of 10,000 rpm for 5 minutes (Wennecke, 2004). Hematocrit was calculated as a ratio of the length of the packed red blood cells to total length of the column. Body mass was measured before and after each test. The two mass measurements were highly positively correlated with each other ( $r = 0.904$ ,  $df = 112$ ,  $P < 0.001$ ), so initial mass was used as a predictor variable when needed in analyses.

## Respirometry

Oxygen consumption ( $VO_2$ ), voluntary wheel running, and home-cage activity were all measured with an 8-cage Promethion multiplex system (Sable Systems International, North Las Vegas, NV, USA), which uses pull-mode flow-through respirometry. A flow generator module pulled a constant air flow multiplexed from the 8 metabolic cages. The gas analysis module measured  $O_2$  consumption while taking changes in water vapour dilution and barometric pressure into account.

Metabolic cages were supplied with a stainless-steel wheel (11.5 cm diameter). Mice had free access to the wheel and air flow around the wheel was unrestricted to integrate wheel activity with respirometry. Wheel revolutions were recorded by a reed switch placed parallel to the wheel, which responded to a magnet placed on the outer wheel rim. Locomotor activity in the cage was measured using the BXZ-1 beam break activity monitor. The activity monitor was designed to record movement across the cage floor and ignore fixed objects such as the hoppers

and wheel. Any cage activity with a speed less than  $0.01 \text{ m s}^{-1}$  was considered as nonlocomotory cage activity (e.g., grooming).

The dwell time (i.e., a period of continuous metabolic and behavioural data monitoring for a given cage) was set at 30 seconds and the inter-leave ratio was set to 4, such that each cage was monitored for 30 sec every 5 minutes. For a period of 24 hours, a total of 288 30-sec  $\text{VO}_2$  measurements were collected for each cage, along with the distance moved on the wheels and in the cage during the same 30-sec as the  $\text{VO}_2$  measurement. Data from the Promethion system were processed and transformed using Expedata (Sable Systems International, North Las Vegas, NV, USA). For each 30-sec measurement, total locomotion speed (in  $\text{m s}^{-1}$ ) was quantified by adding the distance run on the wheel to the distance moved in the cage, then dividing by 30. [note: calculating substitution using only distance moved on the wheel gave similar results (not shown)] Time (in s) spent resting (during the dwell period) was defined as measurements in which the mouse was inactive (absence of locomotor and non-locomotor activity such as grooming, eating, or drinking).

## **Torpor**

For substitution to be accurately assessed, the animal must be normothermic, otherwise the metabolic depression induced by reduced  $T_b$  will return underestimated and/or negative substitution values (see black star in Fig. 1). Despite the choice of the lower temperature condition being predicated on avoiding torpor, inspection of the data suggested that the mice sometimes engaged in torpor during the tests at  $10^\circ\text{C}$ . Torpor is typically identified by significant reductions in  $T_b$ , however it is also distinguishable by periods in which MR is much lower—down to 29% of the expected rate—and followed by a rewarming period in which MR increases by up to 11.6 times (Diedrich et al., 2015). Torpor occurred in bouts during the light cycle at  $10^\circ\text{C}$ . Torpor during these periods was defined by a resting (i.e., inactive in both temperature conditions) mouse displaying lower mean MR at  $10^\circ\text{C}$  than at  $22^\circ\text{C}$  (Fig. 2), during the light cycle (the “daytime”)—as *Peromyscus* only exhibit torpor diurnally (Lynch et al., 1978). Observations where the mouse was engaged in torpor were removed from the raw metabolic measurements before calculation of substitution, which represented 2.7% of all measurements (i.e., 3,586 out of 134,741 observations).

## Tissue sampling

After the third set of tests, mice were sacrificed in groups by a two-step euthanasia process of CO<sub>2</sub> inhalation and cervical dislocation. Post-euthanasia, the body length of the mice was measured from the tip of the snout to the base of the tail using a standard ruler. Tail length was measured separately, and tails were cut at the base and weighed. The skin of the mouse was sprayed with an 80% ethanol solution (Bagchi and Macdougald, 2019). A ventral incision from the mouth to the tail was made and the skin was removed in one piece, excluding the skin on the head, paws, and tail. The fascia and small fat deposits adhering to the skin were kept. The skin was stretched and pinned onto a corkboard and set to dry overnight. An electric razor was used to crop the hair from the skin. The blade was set at 0.4 mm and hair shorter than this length was not cropped. The cropped hair and shaved skin were weighed separately. The area of the skin was measured using a ruler and approximating a rectangle and used as a proxy for the trunk surface area of the mouse.

The heart, kidneys, and liver were removed from the body cavity. Fat deposits, mesentery, and connective tissue were trimmed from the organs, then the organs were rinsed in saline and blotted dry (Morawietz et al., 2004; Scudamore et al., 2014). The organs were all weighed separately. The right gastrocnemius muscle was removed from the mice using a method adapted from Kelly et al. (2017). Fascia was first removed from the muscle complex using a surgical probe (Wang et al., 2017). The calf muscle complex was separated from the tibia-fibula using a surgical probe. The Achilles tendon was then cut midway. The gastrocnemius muscle was separated from the soleus and plantaris muscles using forceps. The muscle was then cut from the condyles of the tibia and fibula, and then weighed.

## Statistical analysis

All models were fitted using ASReml-R 4.0 (Butler et al., 2018). A series of linear models were used to extract individual substitution values separately for each test (see Table S1 for an example). The model was fitted with VO<sub>2</sub> (in mlO<sub>2</sub> s<sup>-1</sup>) as the response variable and the predictors were total activity speed (in m s<sup>-1</sup>), temperature (10 vs 22°C), and their interaction. The reference level for the temperature variable was set at 10°C, such that the model estimate for

the activity speed variable corresponded to the slope of the MR-speed relationship at 10°C, which is equivalent to the  $COA_C$  (in  $mlO_2 m^{-1}$ ). More importantly, the estimate for the “temperature  $\times$  speed” interaction corresponded to how the slope of the MR-speed relationship differed at 22 vs 10°C, which is equivalent to substitution as  $COA_W - COA_C$ . As substitution is describing a difference in the slope of the relationship between metabolic rate (in  $mlO_2 s^{-1}$ ) as function of locomotion speed (in  $m s^{-1}$ ), the units for substitution are in  $mlO_2 m^{-1}$ .

As mice were not forced to run on the wheels, measuring substitution relied on the voluntary activity of the mice. Therefore, tests where mice had less than 5 observations (out of approximately 576) where the speed exceeded  $0.1 m s^{-1}$  in either of the two temperature conditions were removed from the dataset. The speed of  $0.1 m s^{-1}$  was used as a threshold as slower locomotor speeds were not considered to be sufficiently “active” (see Fig. S1). Twelve tests from 11 individual mice were removed. In total, 117 substitution measures were calculated and used for analysis from 46 mice (Table 1).

A quick look at the raw data suggested that the linear models used to extract substitution violated key assumptions (namely homoscedasticity and the normality of residuals). However, the linear models were not used to test for significance, and according to Lande and Arnold (Lande and Arnold, 1983), selection (or substitution) estimates derived from coefficients do not “depend on distribution assumptions”. For completeness, however, substitution was re-estimated using a second method that did not involve model fitting, using the following equation:

$$\text{Eqn 1 } COA = \frac{\Delta VO_2}{\Delta Speed} = \frac{\mu_A - \mu_R}{\mu_S}$$

where  $\mu_A$  represents the mean MR when the mouse was moving at speeds greater than or equal to  $0.1 m s^{-1}$ ,  $\mu_R$  represents the mean MR when the mouse was at rest, and  $\mu_S$  represents the mean speed of the mouse. COA was calculated at 22°C ( $COA_W$ ) and 10°C ( $COA_C$ ) and substitution was calculated as  $COA_W - COA_C$ . This method of deriving substitution will be referred to henceforth as the “second method” (the “first method” being where  $COA_W - COA_C$  is estimated by the “temperature  $\times$  speed” interaction in a linear model, see above and Fig. S2).

Other metabolic parameters—namely daily energy expenditure (DEE) and COA at 10°C and 22°C—were extracted for each mouse for each test. DEE was calculated as the average metabolic rate ( $\text{kcal h}^{-1}$ ) for the test, using the Weir equation (Weir, 1949). COA was extracted from the model (see above). These parameters were extracted for comparison with substitution, along with the body mass and mean locomotor speed for each test for each mouse.

Once the measures of substitution (and other variables like body mass, DEE, and COA) were extracted, the variables were standardized to a mean of 0 and variance of 1 (to ensure that estimates were comparable across studies and improve the interpretability of the results, see Schielzeth, 2010) and linear mixed models (LMM) were run to estimate repeatability. First, “consistency repeatability” (Nakagawa and Schielzeth, 2010) was estimated by only including experimental variables such as test sequence (1 to 3, categorical) and metabolic cage (factor with 8 levels) as fixed effects. Hence, our consistency repeatability estimate accounts for the potential inflation of phenotypic variance caused by non-biological effects such as different measuring devices or “session” effects that were unavoidable in our repeated measures experimental design. The mouse identity was fit as a random effect to estimate the among-individual variance ( $V_I$ ). The within-individual variance was estimated as the residual variance ( $V_e$ ). Statistical significance of  $V_I$  was tested with a likelihood ratio test, which follows an equally weighted mixture of  $\chi^2$ -distributions with one and zero  $df$  ( $\chi^2_{0:1}$ ) (Self and Liang 1987).

While consistency repeatability provides an estimate as to how much variation in substitution is attributable to differences among individuals, it does not account for whether that variation is attributable to individual variation in other traits (such as fixed morphological or physiological traits). It is therefore also desirable to calculate “adjusted repeatability” in substitution after having removed variation caused by other co-varying biological traits (Nakagawa and Schielzeth, 2010). To find possible covariates of substitution, the LMM was re-run with substitution as the dependent variable after including several additional fixed effects and the  $V_I$  and  $V_e$  estimates from this second model were used to calculate “adjusted” repeatability. All of the morphological traits (body mass, body length, tail length, surface area, fur mass, skin mass, tail mass, and organ masses) were included in the model along with torpor bouts (two-level factor, presence, or absence of torpor during a test), hematocrit, resting metabolic rate at 22°C (5<sup>th</sup> percentile of the observations when the mouse was at rest), and the

age of the mouse. Statistical significance of the covariates was tested with a conditional Wald  $F$ -statistic, and the denominator degrees of freedom ( $df_{\text{den}}$ ) were obtained following methods described by Kenward and Roger (1997). The variance inflation factor (VIF)—or the degree to which the variance of a predictor is inflated by correlations with other predictors (Petrakis et al., 1996)—for each fixed effect was calculated to assess whether the model exhibited multicollinearity. Multicollinearity is indicative of a high degree linear dependence between fixed factors—which in turn means that the size of the coefficients and significance in the model are unreliable (Bayman and Dexter, 2021). To amend multicollinearity, fixed effects are often dropped if they are “explained” by another fixed effect (such as body size). The VIFs for all added fixed effects to the model were well below the conservative upper limit of 5 (Akinwande et al., 2015). In fact, all of the morphological and organ mass measurements were weakly correlated, with correlations ranging from -0.28 to 0.29 (Fig. S3). Thus, no fixed effects were removed from the model.

Morphological measurements could not be collected for 10 individuals in the experiment (Table 1)—these mice were omitted from the LMM with the additional fixed effects. Tests where no hematocrit measurements were also excluded. There was one individual with a significantly larger liver mass than the others (2.1g vs  $\text{mean} \pm \text{se} = 0.94 \pm 0.019\text{g}$ ). Although we have no indication of potential disease or infection, we nevertheless removed this individual from the covariate analysis. The total number of mice in this data subset was 35, with 85 observations. For completeness, consistency repeatability was re-calculated on this subset of the data.

### **Caveat: error in raw MR measurements**

The temperature control cabinet distributed air through a central fan on the ceiling of the cabinet, but additional fans were placed on each shelf to homogenise temperature vertically across shelves. However, this had unforeseen consequences on the air flow within the cabinet. As the fans increased the speed of air flowing past the cages adjacent to the fans, air pressure around the cages decreased (the Bernoulli effect, personal communication with Sable Systems, January 18<sup>th</sup>, 2021). Lighton and Halsey (2011) recommend that pull-through systems should be maintained in environments with a slight positive pressure, as this ensures that all excurrent air is captured

by the system. If not all excurrent air is measured, then the flow rate recorded by the system is underestimated, and the MR is in turn underestimated (Lighton, 2008). Overall, this created an experimental setup in which raw measurements of MR were different between cages, with underestimated MR measurements from cages closest to the fans. One month before this study, the same system and the same mice were used in an experiment where metabolic cages were maintained outside the temperature control cabinet. In that experiment there was no reported effect of cage on MR, and DEE was highly repeatable at 0.786 (Abdeen et al., 2022). Variation in air pressure around the cages was a major source of measurement error in our experiment, increasing the residual variance and resulting in low repeatability estimates for DEE. The extent to which measurement error in raw MR measurements affected repeatability of substitution is unknown, but is probably less problematic than for DEE because substitution was calculated as the difference in the COA at 10 vs 22°C, which were presumably equally underestimated (both measured within a test with the same cage, see Fig. 3C, and mice were assigned to a different cage on each test).

## RESULTS

### Repeatability

Body mass had the highest repeatability ( $R = 0.90$ ), followed by mean locomotor speed ( $R = 0.648$  at 10°C and  $R = 0.491$  at 22°C). Although COA was significantly repeatable at 22°C ( $R = 0.269$ ), repeatability was lower and nonsignificant at 10°C ( $R = 0.126$ ,  $P = 0.143$ ). The repeatability of DEE in both temperatures was low and nonsignificant (Table 2).

### Repeatability of substitution

Overall, the average substitution was  $0.0168 \text{ mL O}_2 \text{ m}^{-1}$  (Table 1). Consistency repeatability of substitution was  $R \pm \text{se} = 0.306 \pm 0.134$  ( $P = 0.003$ ), indicating that approximately one third of variation in substitution was attributable to differences among individuals (Table 3A, Fig. 4A). The metabolic cage and test sequence did not significantly influence substitution (results not shown). Results remained qualitatively similar when using the second method for deriving substitution (Table 3B, Fig. 4B) with a consistency repeatability ( $\pm \text{se}$ ) of  $R \pm \text{se} = 0.279 \pm 0.133$  ( $P =$

0.006). There was a very strong positive correlation between substitution estimates derived from the two methods ( $r = 0.952$ ,  $P < 0.001$ ).

## Covariates

Adjusted repeatability (calculated using  $V_I$  and  $V_e$  from a LMM with multiple fixed effects, see Table 4) was  $R_{\pm se} = 0.074 \pm 0.112$  ( $P = 0.237$ ; Table 3C). The reduction in repeatability was not due to the use of a slightly different subset of the data (due to missing values for the covariates), because re-calculating consistency repeatability on the subset of the “complete cases” data yielded an estimate of  $R_{\pm se} = 0.220 \pm 0.121$  ( $P = 0.016$ ; Table 3D), which is close to the one obtained above when using all observations (Table 3A). The same significant covariates were found using substitution as derived from the second method (results not shown). Body length was significantly and positively related to substitution (Table 4, Fig. 5A). The trunk surface area of the mouse was significantly and negatively related to substitution (Table 4, Fig. 5B). Finally, heart mass was significantly and positively related to substitution (Table 4, Fig. 5C). Conclusions about covariates remained the same change when analysing substitution calculated with the second method (Table S2).

## DISCUSSION

Kemp (2006) considered endothermy as a paradigm for the evolution of complex traits. Indeed, endothermy has multifarious implications for many complex traits like  $T_b$ , metabolic rate, and locomotor activity. Substitution is a complex trait that emerges from interplay between metabolic rate, locomotor activity, and heat dissipation. Complex traits are influenced by several genetic and environmental factors, and as such individual differences in complex traits should be repeatable over time. While it was already known that substitution is influenced by the environment (McNamara et al., 2004; Travis et al., 1999), the individual repeatability of substitution remained unknown. This study established that around one third of the phenotypic variation in substitution is attributable to individual differences. Moreover, the expression of substitution is dependant on variation in underlying sub-organismal traits—in this case, morphology of the organism. Given that substitution is complex and its expression is almost

entirely dependant on other traits, understanding how it evolves requires consideration of the physiological roles of both the significant (and non-significant) covariates.

Conductivity of fur, skin, and fat determines how much metabolic heat lost to peripheral tissue is retained vs. lost to the outer environment (González-Alonso, 2012), and it is therefore naturally expected that conductivity has a strong influence on substitution. Here, we could not directly measure individual differences in conductivity, which is best quantified as the heat transfer and temperature difference across the gradient of the tissue (Boyles and Bakken, 2007; Jacobsen, 1980; Knight, 1987). However, we measured the mass of the skin and pelage, which have been used as a validated proxy for conductivity (Barnett, 1959). Despite the supposedly strong influence that skin and fur mass have on conductance, we found no relationship with substitution. Peripheral fat deposits are also known to influence conductivity, however *Peromyscus* mice do not accumulate much subcutaneous fat and remain lean throughout their lifetime (CMM, pers. obs.; fat deposits were too small to excise and measure accurately).

A strong and negative relationship was found between the trunk surface area of the mice and substitution (Fig. 5B). All else being equal, heat transfer is directly proportional to surface area through which heat is being conducted. Indeed, it has been well established that the ratio of surface area to mass changes the rate of heat exchange: a greater surface area to mass ratio results in more dry heat loss in an animal (Mitchell et al., 2018). In mice, the bulk of dry heat loss is facilitated by trunk surface area—which probably explains why surface area strongly and negatively covaried with substitution. It is worth mentioning that rodents rely on dry heat exchange to avoid (often lethal) hyperthermia (Rezende and Bacigalupe, 2015).

Intuitively, a mouse with a longer body should experience higher heat loss than a shorter one, as the surface area would be greater for a longer body—but instead body length was positively related to substitution (Fig. 5A). We currently have no clear explanation for this finding, but one possibility is that the body length effect is attributable to variation in head size, which was not directly measured in this study but was included in overall body length. In *Peromyscus leucopus*, the length of the skull is 26% of the total body length and coefficient of variation for body length and skull length are 4.19 and 2.27, respectively (Clark, 1941). The brains of mammals are more sensitive to temperature changes and remain cooler than the rest of the body (Matsuda-Nakamura and Nagashima, 2014). Due to this sensitivity, the brain relies on

heat loss mechanisms beyond dry heat transfer. Selective brain cooling works through several mechanisms that are often species dependant, but two mechanisms that are found in most mammals is the pre-cooling of arterial blood headed to the brain and the drainage of cooled venous blood to veins around the brain (Caputa, 2004). Thus, there is a smaller temperature gradient between the environment and the surface of the head, resulting in less heat loss in the cold. Hence, mice with a bigger head (and a longer body) would lose proportionally less heat than mice with a smaller head (and a shorter body), and this reduced heat loss might have contributed to greater substitution. We recognise, however, that this explanation rests on many assumptions that will have to be verified in future.

A significant positive relationship was found between heart mass and substitution (Fig. 5C). A larger heart may be an advantage to exercising mice experiencing cold temperatures. Cold temperatures induce cardiovascular changes—not only does vasoconstriction reduce blood flow to peripheral tissue, but arteries supplying the extremities decrease their flow and conductance (up to 40% for the femoral artery, González-Alonso, 2012). Reducing blood flow directly causes less heat loss in peripheral tissue and in the extremities. However, there are consequences to reducing blood flow. As blood volume is concentrated in core tissues, blood pressure increases accordingly, which in turn overloads the heart (Choo et al., 2018; Halonen et al., 2011). This pressure overload results in cardiac hypertrophy and contractile abnormalities, which can result in heart failure (Lu and Xu, 2013). A naturally larger heart can accommodate increases in blood pressure better than a smaller heart (as it has the capacity to safely hold greater blood volume). Therefore, a likely explanation for the effect of heart mass on substitution is that mice with larger hearts lose less heat from active skeletal muscles than mice with smaller hearts due to greater reductions in peripheral blood flow and arterial conductance.

## Experimental constraints

It was necessary to remove torpor bouts from the dataset, as substitution can only be correctly estimated when the animal is normothermic (Fig. 1). In this study, mice used both substitution and torpor in conjunction at different times in the light cycle—which Geiser (2020) suggested is common for *Peromyscus*. White-footed mice have been well-established to engage in torpor when faced with cold ambient temperature (Lynch et al., 1978; Lynch et al., 1978; Rhodes,

1980) or food deprivation (Diedrich et al., 2015). Torpor has been shown to be induced at temperatures as high as 15°C, but only in laboratory settings where movement was severely restricted (Hill, 1975). In wild populations, torpor is more typically observed at lower temperatures (Lynch et al., 1978a; Lynch et al., 1978b). Shorter photoperiods (less than 12 hours light) have been positively correlated with torpor bouts, but members of *Peromyscus* have been described as using daily torpor year-round whenever the ambient temperature falls (Geiser, 2020). No relationship was found between the use of torpor (as identified by torpor bouts present during tests in individuals) and substitution. However, torpor is not an “all-or-nothing” phenomenon as quantified here, and future studies should include concomitant  $T_b$  monitoring to quantify individual variation in torpor use and potential covariation with substitution. In deer mice (*Peromyscus maniculatus*), some individuals are “torpor sensitive” while others are “torpor resistant” (Sheafor and Snyder, 1996). Use of a lower sub-thermoneutral temperature condition, potentially in conjunction with food restriction, would induce spontaneous torpor in a higher proportion of the population, thereby facilitating a comprehensive study on individual (co)variation in substitution and torpor use (Lynch et al., 1978b).

We decided to use 22°C for the warm temperature treatment because it corresponded to the housing conditions of the mice. Compared with mice housed within their TNZ (between 27.8-34.5°C; (Deavers and Hudson, 1981), mice housed at 20-22°C do exhibit signs of cold stress (namely impaired immune function and increased thermogenesis; (Deavers and Hudson, 1981)). Although cold stress at 22°C can be mitigated by providing food and nesting material, we also decided to provide no nesting material during metabolic measurements. The main implication of these ( $T_a$  lower than TNZ and absence of nesting material) is that we did not quantify the full extent of substitution. Future research should include a temperature in the TNZ, but it is also possible that mice would be averse to activity at a temperature within the TNZ, as even at these temperatures rodents risk hyperthermia (Rezende and Bacigalupe, 2015; Speakman and Król, 2010; van Klinken et al., 2013). This would make measurements of substitution based on voluntary exercise more difficult because mice may not express the full range of activity intensity in the TNZ.

The use of substitution may shift with acclimation to cold, in turn changing its degree of utilization within individuals. Indeed, cold acclimatization increases the capacity for non-shivering thermogenesis (Van Sant and Hammond, 2008) with paired changes in physiological traits and organ size (Hayward et al., 2022; Nedergaard and Cannon, 2013). The mice in this experiment were never acclimated to the lower temperature condition, as the length of time they remained exposed (2 days) was significantly shorter than other experimental “short-term” acclimatization periods used (e.g., 25 days, Andrew et al., 2019). *Peromyscus* do not adjust their thermal conductance (through fur) seasonally (Boyles and Bakken, 2007; Hayward et al., 2022), such that our conclusions regarding the role of conductivity and substitution are likely to remain the same in cold acclimated mice. Still, an interesting avenue for future research would be to compare the use of substitution in cold acclimated vs non-acclimated groups of mice, and determine if the relationship between substitution and organ size changes with cold acclimatization.

## Conclusion

Our study provides the first repeatability estimate for substitution. Moreover, we identified key morphological traits underlying individual variation in the substitution. However, given the constraints of the experimental set-up and the use of torpor, it is likely that we have not identified the full extent of individual variation in substitution. We suggest future research should focus on mitigating these constraints and expand upon this work by measuring substitution in conjunction with  $T_b$  monitoring over a wider range of  $T_a$ . Subsequent work should also be done to illuminate whether substitution conveys direct fitness advantages and/or is involved in a trade-off with necessary heat-loss. Indeed, traits associated with increased substitution might increase the risks of hyperthermia at higher  $T_a$ , and the optimal level of substitution vs capacity for heat dissipation might depend on the thermal regime.

## Acknowledgements

We thank Paul Agnani for assisting with set-up of experimental procedures and technical support. We also thank the University of Ottawa Animal Care and Veterinary Service for

maintaining the study population, technical training, and facilities and material for procedures. Finally, a special thanks to Mathieu Videlier for providing feedback on statistical analyses, and two anonymous reviewers for their constructive criticism on an earlier version of the manuscript.

### Competing Interests

The authors declare no competing or financial interests.

### Funding

This research was supported by Natural Sciences and Engineering Research Council of Canada (NSERC; Grant no. 2016-04418) and Canada Research Chair programs.

### Data Availability

Dataset and R code available on figshare (<https://doi.org/10.6084/m9.figshare.20401944.v1>).

### References

- Abdeen, A., Agnani, P. and Careau, V.** (2022). The active mouse rests within: Energy management among and within individuals. *Functional Ecology* 1365-2435.13979.
- Akinwande, M. O., Dikko, H. G. and Samson, A.** (2015). Variance Inflation Factor: As a Condition for the Inclusion of Suppressor Variable(s) in Regression Analysis. *Open Journal of Statistics* **05**, 754–767.
- Andrew, J. R., Garland, T., Chappell, M. A., Zhao, M. and Saltzman, W.** (2019). Effects of short- and long-term cold acclimation on morphology, physiology, and exercise performance of California mice (*Peromyscus californicus*): potential modulation by fatherhood. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology* **189**, 471–487.
- Araya-Ajoy, Y. G., Mathot, K. J. and Dingemanse, N. J.** (2015). An approach to estimate short-term, long-term and reaction norm repeatability. *Methods in Ecology and Evolution* **6**, 1462–1473.

- Arnold, W., Ruf, T. and Kuntz, R.** (2006). Seasonal adjustment of energy budget in a large wild mammal, the Przewalski horse (*Equus ferus przewalskii*). II. Energy expenditure. *Journal of Experimental Biology* **209**, 4566–4573.
- Bagchi, D. P. and Macdougald, O. A.** (2019). Identification and dissection of diverse mouse adipose depots. *Journal of Visualized Experiments* **2019**, 1–9.
- Barnett, S. A.** (1959). The skin and hair of mice living at a low environmental temperature. *Quarterly Journal of Experimental Physiology and Cognate Medical Sciences* **44**, 35–42.
- Bayman, E. O. and Dexter, F.** (2021). Multicollinearity in Logistic Regression Models. *Anesthesia & Analgesia* **133**, 362–365.
- Boake, C. R. B.** (1989). Repeatability: Its role in evolutionary studies of mating behavior. *Evol Ecol* **3**, 173–182.
- Borniger, J. C. and Nelson, R. J.** (2017). Photoperiodic regulation of behavior: *Peromyscus* as a model system. *Seminars in Cell and Developmental Biology* **61**, 82–91.
- Boyles, J. G. and Bakken, G. S.** (2007). Seasonal changes and wind dependence of thermal conductance in dorsal fur from two small mammal species (*Peromyscus leucopus* and *Microtus pennsylvanicus*). *Journal of Thermal Biology* **32**, 383–387.
- Bryant, D. M., Hails, C. J. and Prys-Jones, R.** (1985). Energy Expenditure by Free-Living Dipper (Cinclus cinclus) in Winter. *The Condor* **87**, 177.
- Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J. and Thompson, R.** (2018). ASReml-R Reference Manual Version 4. *ASReml-R Reference Manual* 176.
- Caputa, M.** (2004). Selective brain cooling: a multiple regulatory mechanism. *Journal of Thermal Biology* **29**, 691–702.
- Chappell, M. A. and Hammond, K. A.** (2004). Maximal aerobic performance of deer mice in combined cold and exercise challenges. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology*.
- Chappell, M. A., Garland, T., Rezende, E. L. and Gomes, F. R.** (2004). Voluntary running in deer mice: Speed, distance, energy costs and temperature effects. *Journal of Experimental Biology* **207**, 3839–3854.
- Clark, F. H.** (1941). Correlation and Body Proportions in Mature Mice of the Genus *Peromyscus*. *Genetics* **26**, 283.
- Deavers, D. R. and Hudson, J. W.** (1981). Temperature Regulation in Two Rodents (*Clethrionomys gapperi* and *Peromyscus leucopus*) and a Shrew (*Blarina brevicauda*) Inhabiting the Same Environment. *Physiological Zoology* **54**, 94–108.

- Diedrich, V., Kumstel, S. and Steinlechner, S.** (2015). Spontaneous daily torpor and fasting-induced torpor in Djungarian hamsters are characterized by distinct patterns of metabolic rate. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology*.
- Geiser, F.** (2020). Seasonal Expression of Avian and Mammalian Daily Torpor and Hibernation: Not a Simple Summer-Winter Affair†. *Frontiers in Physiology* **11**, 1–19.
- González-Alonso, J.** (2012). Human thermoregulation and the cardiovascular system. *Experimental Physiology* **97**, 340–346.
- Hart, J. S.** (1950). *Interrelations of Daily Metabolic Cycle, Activity, and Environmental Temperature of Mice*.
- Hart, J. S. and Heroux, O.** (1955). Exercise and temperature regulation in lemmings and rabbits. *Canadian Journal of Biochemistry and Physiology*.
- Hayes, J. P. and Jenkins, S. H.** (1997). Individual Variation in Mammals. *Journal of Mammalogy* **78**, 274–293.
- Hayward, L., Robertson, C. E. and McClelland, G. B.** (2022). Phenotypic plasticity to chronic cold exposure in two species of *Peromyscus* from different environments. *J Comp Physiol B* **192**, 335–348.
- Hedrick, M. S. and Hillman, S. S.** (2016). What drove the evolution of endothermy? *Journal of Experimental Biology* **219**, 300–301.
- Hetem, R. S., Maloney, S. K., Fuller, A. and Mitchell, D.** (2016). Heterothermy in large mammals: Inevitable or implemented? *Biological Reviews* **91**, 187–205.
- Hill, R. W.** (1975). Daily torpor in *Peromyscus leucopus* on an adequate diet. *Comparative Biochemistry and Physiology – Part A: Physiology* **51**, 413–423.
- Humphries, M. M. and Careau, V.** (2011). Heat for nothing or activity for free? Evidence and implications of activity-thermoregulatory heat substitution. In *Integrative and Comparative Biology*, pp. 419–431.
- Jacobsen, N. K.** (1980). Differences of thermal properties of white-tailed deer pelage between seasons and body regions. *Journal of Thermal Biology* **5**, 151–158.
- Kelly, S. A., Gomes, F. R., Kolb, E. M., Malisch, J. L. and Garland, T.** (2017). Effects of activity, genetic selection and their interaction on muscle metabolic capacities and organ masses in mice. *Journal of Experimental Biology* **220**, 1038–1047.
- Kemp, T. S.** (2006). The origin of mammalian endothermy: a paradigm for the evolution of complex biological structure. *Zoological Journal of the Linnean Society* **147**, 473–488.

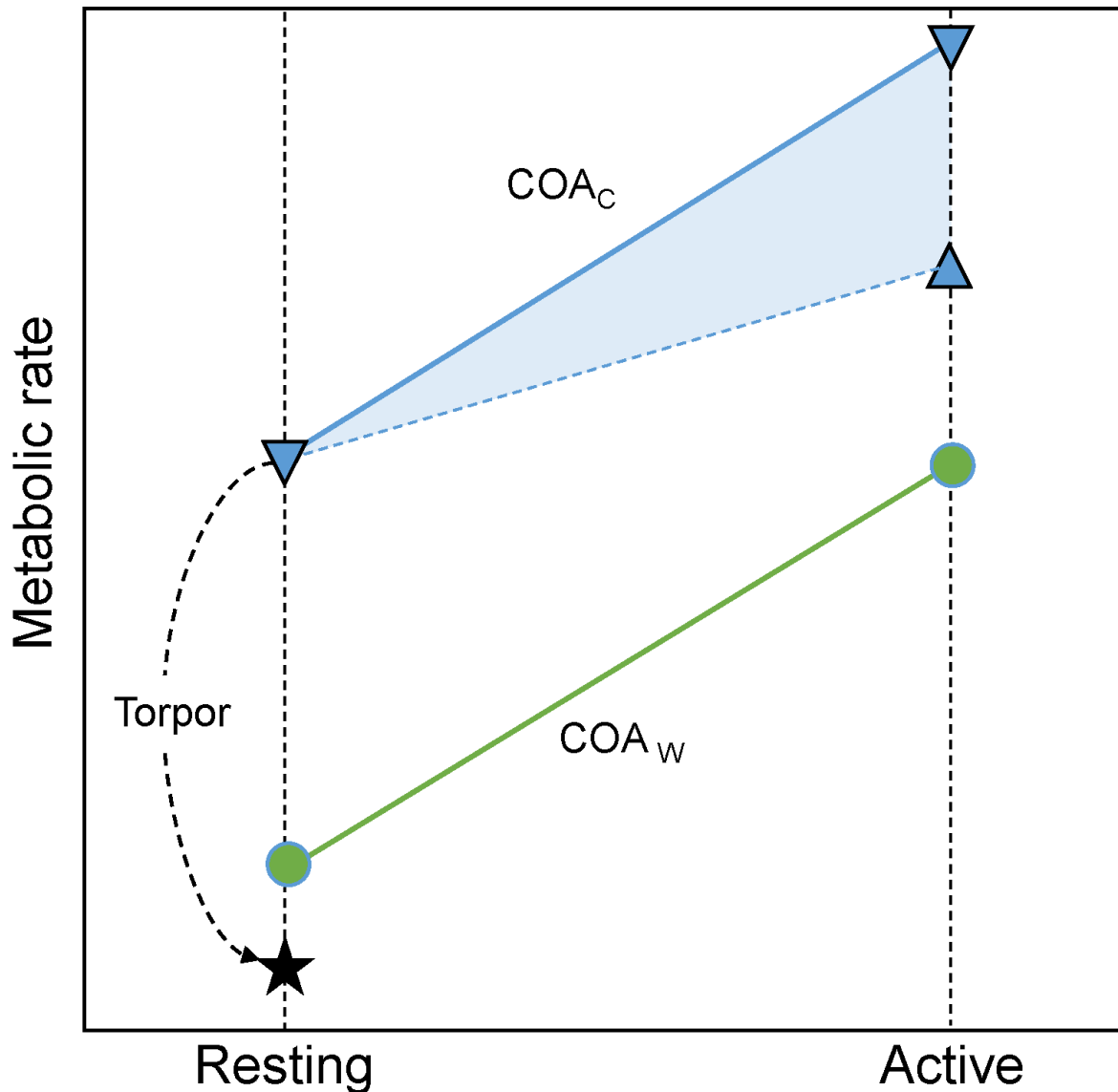
- Kenward, M. G. and Roger, J. H.** (1997). Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood. *Biometrics* **53**, 983.
- Knight, F. M.** (1987). The Development of Pelage Insulation and Its Relationship to Homeothermic Ability in an Altricial Rodent, *Peromyscus leucopus*. *Physiological Zoology* **60**, 181–190.
- Konarzewski, M. and Diamond, J.** (1995). Evolution of basal metabolic rate and organ masses in laboratory mice. *Evolution* **49**, 1239–1248.
- Lande, R. and Arnold, S. J.** (1983). The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226.
- Lefèvre, J. and Auget, A.** (1931). Le problème des relations entre les chaleurs du travail et du repos. Solution et lois. Pourquoi le travail est plus économique aux basses températures. *Comptes rendus hebdomadaires des séances de l'Académie des sciences* **192**, 818–820.
- Levesque, D. L., Nowack, J. and Stawski, C.** (2016). Modelling mammalian energetics: the heterothermy problem. *Climate Change Responses* **3**,.
- Lighton, J. R. B.** (2008). *Measuring Metabolic Rates: A Manual for Scientists*.
- Lighton, J. R. B. and Halsey, L. G.** (2011). Flow-through respirometry applied to chamber systems: Pros and cons, hints and tips. *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology* **158**, 265–275.
- Liwanag, H. E. M., Williams, T. M., Costa, D. P., Kanatous, S. B., Davis, R. W. and Boyd, I. L.** (2009). The effects of water temperature on the energetic costs of juvenile and adult California sea lions (*Zalophus californianus*): The importance of skeletal muscle thermogenesis for thermal balance. *Journal of Experimental Biology* **212**, 3977–3984.
- Lynch, G. R., White, S. E., Grundel, R. and Berger, M. S.** (1978). Effects of photoperiod, melatonin administration and thyroid block on spontaneous daily torpor and temperature regulation in the white-footed mouse, *Peromyscus leucopus*. *Journal of Comparative Physiology B* **125**, 157–163.
- Maresh, J. L., Adachi, T., Takahashi, A., Naito, Y., Crocker, D. E., Horning, M., Williams, T. M. and Costa, D. P.** (2015). Summing the strokes: Energy economy in northern elephant seals during large-scale foraging migrations. *Movement Ecology* **3**,.
- Matsuda-Nakamura, M. and Nagashima, K.** (2014). Protection of the brain against heat damage. *The Journal of Physical Fitness and Sports Medicine* **3**, 217–221.
- McNamara, J. M., Ekman, J. and Houston, A. I.** (2004). The effect of thermoregulatory substitution on optimal energy reserves of small birds in winter. *Oikos* **105**, 192–196.

- Menzies, A. K., Studd, E. K., Majchrzak, Y. N., Peers, M. J. L. and Boutin, S.** (2020). Body temperature, heart rate, and activity patterns of two boreal homeotherms in winter: homeostasis, allostasis, and ecological coexistence. *Functional Ecology* **1**.
- Mitchell, S. E., Tang, Z. H., Kerbois, C., Delville, C., Deros, D., Green, C. L., Wang, Y., Han, J. J. D., Chen, L., Douglas, A., et al.** (2017a). The effects of graded levels of calorie restriction: VIII. Impact of short term calorie and protein restriction on basal metabolic rate in the C57BL/6 mouse. *Oncotarget* **8**, 17453–17474.
- Mitchell, G., Sittert, S. van, Roberts, D. and Mitchell, D.** (2017b). Body surface area and thermoregulation in giraffes. *Journal of Arid Environments* **145**, 35–42.
- Mitchell, D., Snelling, E. P., Hetem, R. S., Maloney, S. K., Strauss, W. M. and Fuller, A.** (2018). Revisiting concepts of thermal physiology: Predicting responses of mammals to climate change. *J Anim Ecol* **87**, 956–973.
- Morawietz, G., Ruehl-Fehlert, C., Kittel, B., Bube, A., Keane, K., Halm, S., Heuser, A. and Hellmann, J.** (2004). Revised guides for organ sampling and trimming in rats and mice – Part 3. *Experimental and Toxicologic Pathology* **55**, 433–449.
- Nedergaard, J. and Cannon, B.** (2013). UCP1 mRNA does not produce heat. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1831**, 943–949.
- Pauls, R. W.** (1981). *Energetics of the red squirrel: A laboratory study of the effects of temperature, seasonal acclimatization, use of the nest and exercise.*
- Petratis, P. S., Dunham, A. E. and Niewiarowski, P. H.** (1996). Inferring Multiple Causality: The Limitations of Path Analysis. *Functional Ecology* **10**, 421.
- Pohl, H. and West, G. C.** (1973). Daily and seasonal variation in metabolic response to cold during rest and forced exercise in the common redpoll. *Comparative Biochemistry and Physiology – Part A: Physiology* **45**, 851–867.
- Raichlen, D. A., Gordon, A. D., Muchlinski, M. N. and Snodgrass, J. J.** (2010). Causes and significance of variation in mammalian basal metabolism. *J Comp Physiol B* **180**, 301–311.
- Rezende, E. L. and Bacigalupe, L. D.** (2015). Thermoregulation in endotherms: physiological principles and ecological consequences. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology* **185**, 709–727.
- Rhodes, D. H.** (1980). Effects of temperature, photoperiod, and daily torpor on thyroid function in the white-footed mouse, *Peromyscus leucopus*. *General and Comparative Endocrinology* **42**, 134–140.
- Roche, D. G., Careau, V. and Binning, S. A.** (2016). Demystifying animal ‘personality’ (or not): why individual variation matters to experimental biologists. *Journal of Experimental Biology* jeb.146712.

- Santos, C. D., Cramer, J. F., Pârâu, L. G., Miranda, A. C., Wikelski, M. and Dechmann, D. K. N.** (2015). Personality and morphological traits affect pigeon survival from raptor attacks. *Sci Rep* **5**, 15490.
- Self, S. C. and K. Y. Liang.** (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. *J. Am. Stat. Soc.* **82**, 605-610.
- Schielzeth, H.** (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution* **1**, 103–113.
- Scudamore, C. L., Busk, N. and Vowell, K.** (2014). A simplified necropsy technique for mice: Making the most of unscheduled deaths. *Laboratory Animals* **48**, 342–344.
- Sheafor, B. A. and Snyder, G. K.** (1996). Energy partitioning in torpor-sensitive and torpor-resistant deer mice ( *Peromyscus maniculatus* ). *Can. J. Zool.* **74**, 1179–1186.
- Škop, V., Liu, N., Guo, J., Gavrilova, O. and Reitman, M. L.** (2020). The contribution of the mouse tail to thermoregulation is modest. *American Journal of Physiology - Endocrinology and Metabolism* **319**, F438–F446.
- 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. *Journal of Animal Ecology* no-no.
- Stearns, S. C.** (2000). Life history evolution: successes, limitations, and prospects. *Naturwissenschaften* **87**, 476–486.
- Travis, J., McManus, M. G., Baer, C. F., Fisheries, F. and Collins, F.** (1999). Sources of variation in physiological phenotypes and their evolutionary significance. *American Zoologist* **433**, 422–433.
- van Klinken, J. B., Berg, S. A. A. van den and Dijkl, K. W. van** (2013). Practical aspects of estimating energy components in rodents. *Frontiers in Physiology* **4 MAY**,.
- Van Sant, M. J. and Hammond, K. A.** (2008). Contribution of Shivering and Nonshivering Thermogenesis to Thermogenic Capacity for the Deer Mouse ( *Peromyscus maniculatus* ). *Physiological and Biochemical Zoology* **81**, 605–611.
- Wang, C., Yue, F. and Kuang, S.** (2017). Muscle Histology Characterization Using H&E Staining and Muscle Fiber Type Classification Using Immunofluorescence Staining. *Bio-Protocol* **7**,.
- Weir, J. B. de V.** (1949). New methods for calculating metabolic rate with special reference to protein metabolism. *The Journal of Physiology* **109**, 1–9.
- Wennecke, G.** (2004). Hematocrit - a review of different analytical methods. *Acutecaretesting.Org* 1–9.

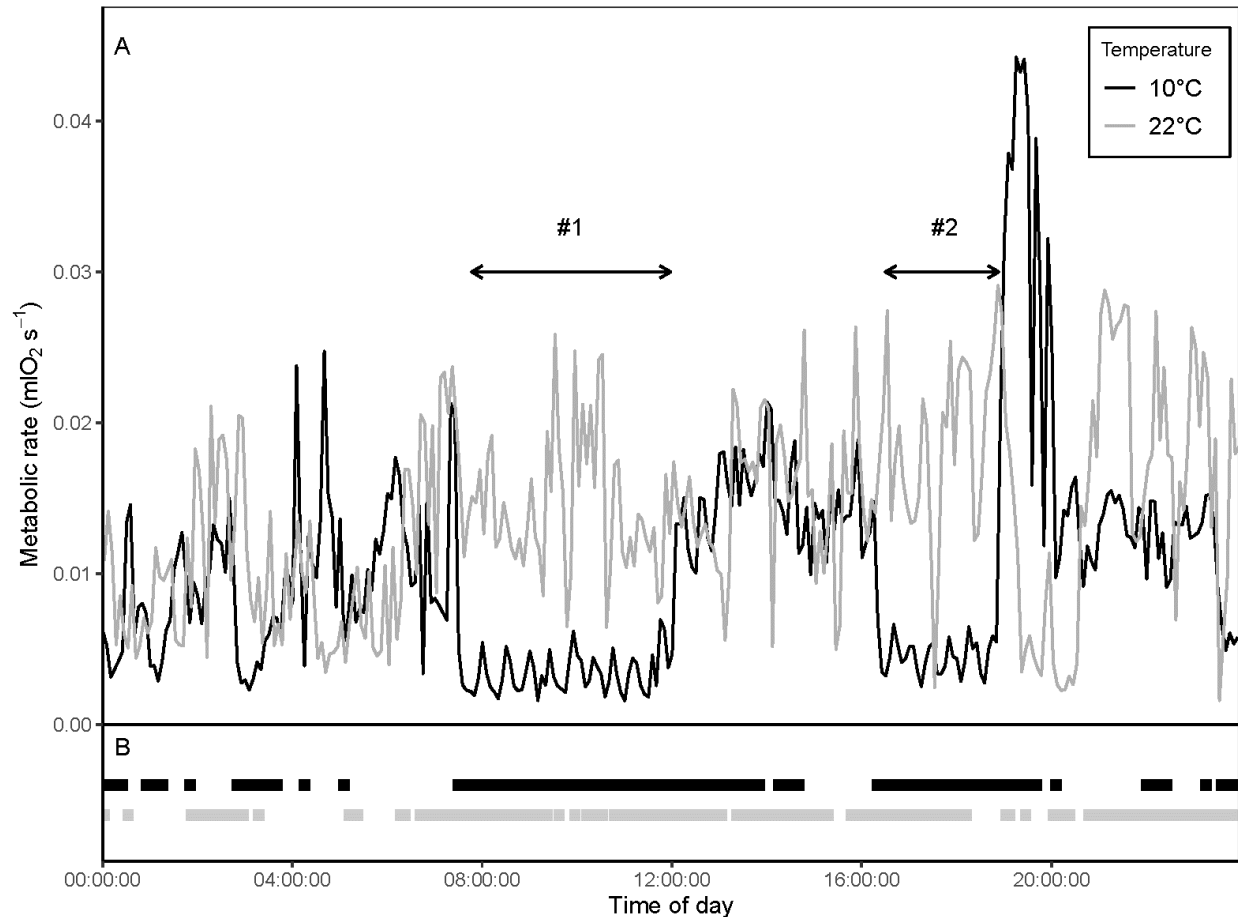
- Wilson, A. J.** (2008). Why  $h^2$  does not always equal  $VA/VP$ ? Why  $h^2$  does not always equal  $VA/VP$ ? *Journal of Evolutionary Biology* **21**, 647–650.
- Wilson, A. J.** (2018). How should we interpret estimates of individual repeatability? *Evolution Letters* **2**, 4–8.
- Wolak, M. E., Fairbairn, D. J. and Paulsen, Y. R.** (2012). Guidelines for estimating repeatability. *Methods in Ecology and Evolution* **3**, 129–137.
- Yousef, M. K., Robertson, W. D., Dill, D. B. and Johnson, H. D.** (1973). Energetic cost of running in the antelope ground squirrel *Ammospermophilus leucurus*. *Physiological Zoology* **46**, 139–147.

## Figures and Tables

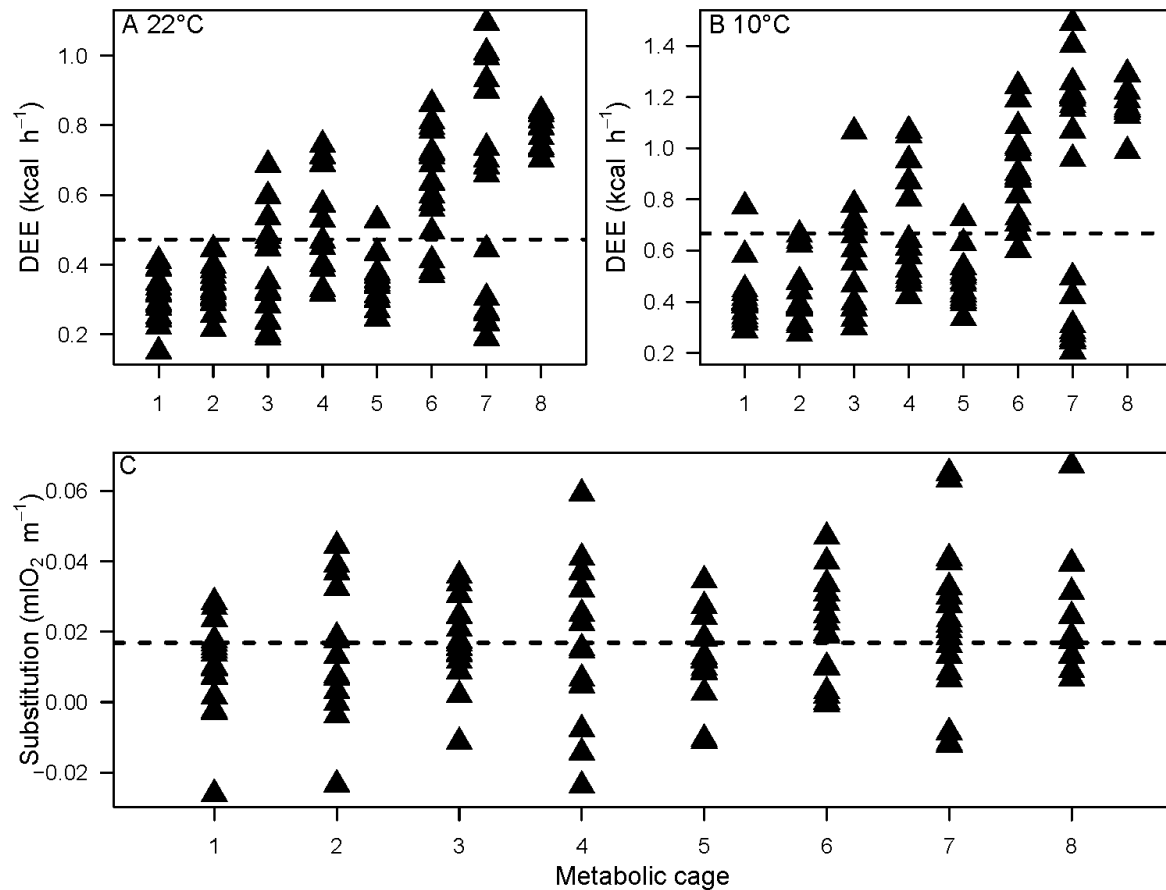


**Fig. 1. Hypothetical relationship between metabolic rate (MR) as function of activity intensity in an endotherm.** Substitution can be quantified as the difference between cost of activity (COA) at a warm temperature ( $COA_W$ , green line connecting dots) and a substantially colder temperature ( $COA_C$ , blue lines connecting triangles). When  $COA_W$  and  $COA_C$  are parallel (solid lines), there is no substitution and the costs of thermoregulation and activity are 100% additive. When the  $COA_C$  is shallower than  $COA_W$  (solid green line vs dotted blue line), there is

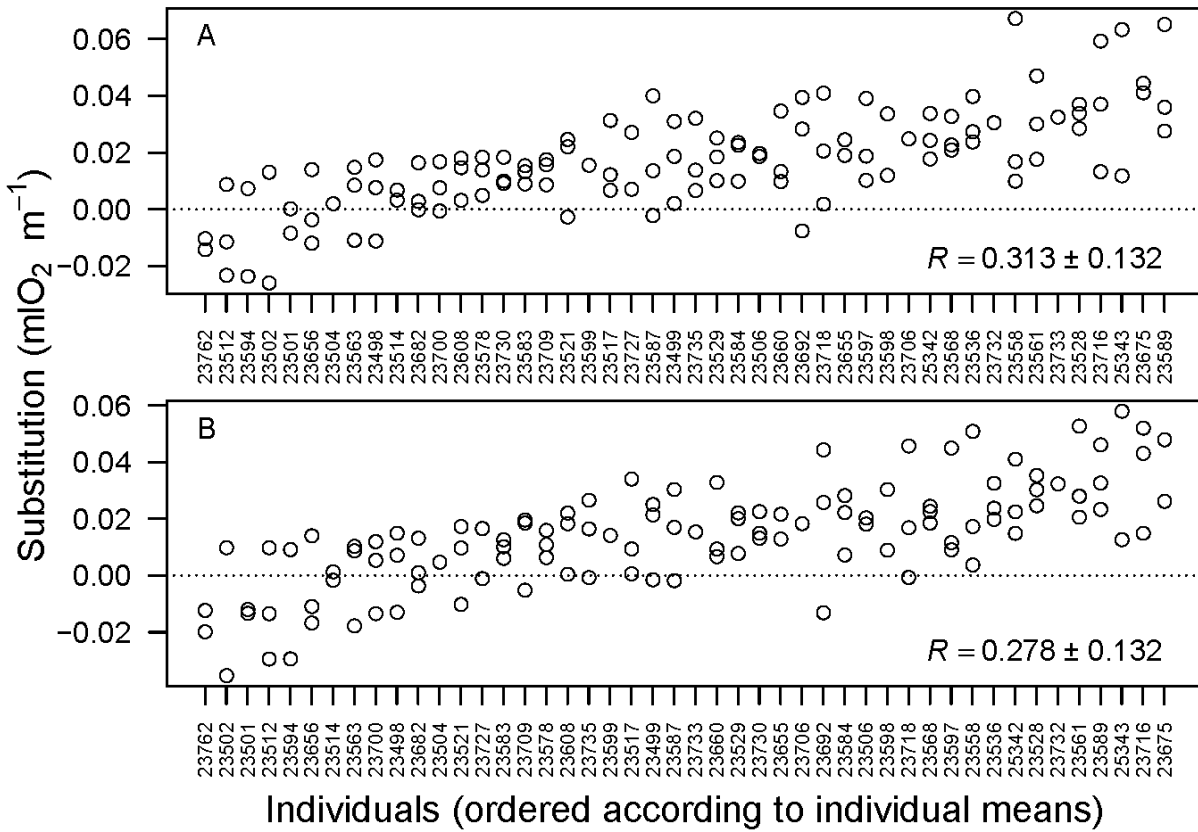
substitution, and the shaded area represents net energy savings from substitution. In some circumstances, a resting animal in the cold might enter torpor (black star) and therefore reduce the estimated cost of resting below the cost of resting estimated for higher ambient temperature, which would return a negative substitution estimate. In this study, all substitution estimates were calculated after excluding raw metabolic measurements presumably made on torpid animals.



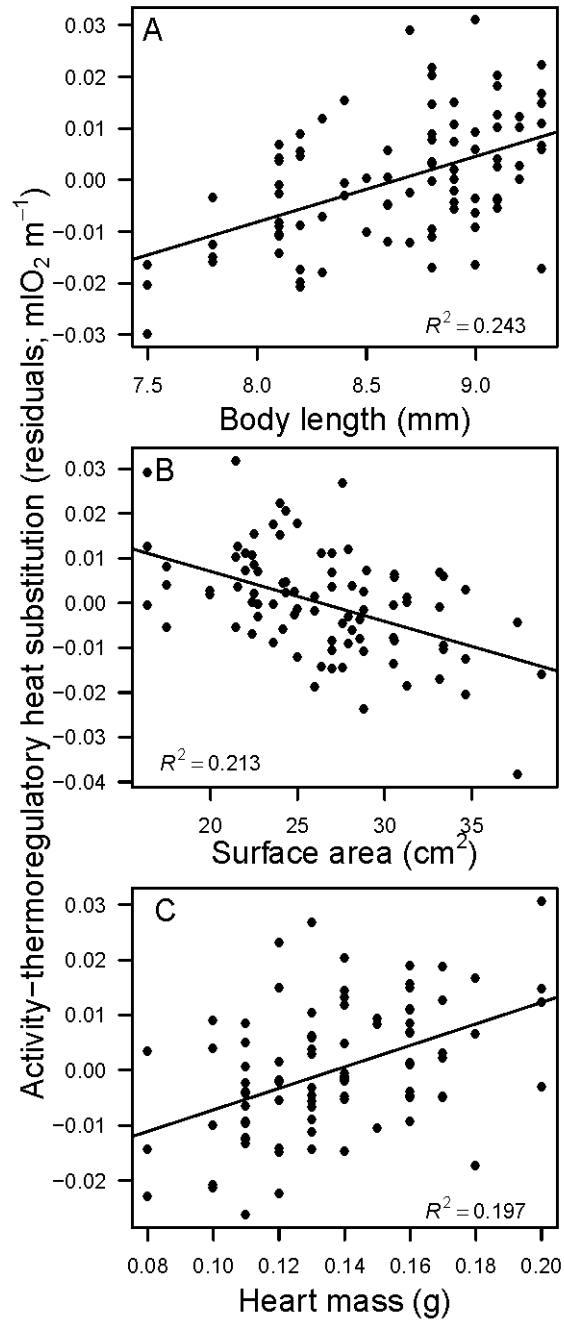
**Fig. 2. Identification of torpor bouts in raw metabolic measurements.** A) Oxygen consumption (mL s<sup>-1</sup>) and B) inactivity (solid bars) as function of time of day (hours) for a single individual (mouse #23512, test #1), at 10°C (black line and bars) and 22°C (grey line and bars). Time has been limited to one 24-hour period within the test for both temperature conditions to better illustrate torpor bouts #1 (8:00 to 12:00) and #2 (16:00 to 19:00), where oxygen consumption for the same animal at rest was lower at 10°C than 22°C for the entire duration of the torpor bout. Also note how torpor bout #2 is immediately followed by a marked increased in oxygen consumption despite inactivity, which is presumably reflects the re-warming phase.



**Fig. 3. Cage effects in raw metabolic measurements but not in substitution.** Daily energy expenditure (DEE, kcal h<sup>-1</sup>) at A) 22°C and B) 10°C, and C) substitution (mlO<sub>2</sub> m<sup>-1</sup>) as a function of metabolic cage (1 through 8). Data points are used to indicate individual tests and the dashed line indicate the population average in each graph. Note that for DEE, cage effects are similar at 10°C (A) and 22°C (B), but absent in substitution measurements (C).



**Fig. 4. Repeatability in activity-thermoregulatory heat substitution.** Substitution derived from A) model coefficients (see Table 1) and B) the second method (see text) in 46 white-footed mice, ordered on the x-axis from lowest to highest average, showing the relative importance of among- and within-individual variance. Consistency repeatability estimate is indicated for substitution ( $R \pm SE$ )



**Fig. 5. Covariates underlying individual variation in substitution.** Partial residuals of substitution (mLO<sub>2</sub> m<sup>-1</sup>) derived from the full linear mixed model as a function of A) body length (mm), B) surface area (cm<sup>2</sup>), and c) heart mass (g) in 35 white-footed mice (for statistical significance, see Table 4). Solid lines indicate the line of best fit. Also shown are the adjusted  $R^2$  for each covariate.

**Table 1.** Descriptive statistics, including units, number of measurements ( $n_{\text{obs}}$ ), number of individuals measured ( $N_{\text{ID}}$ ), mean, standard deviation (SD), and range, locomotor activity, daily energy expenditure (DEE), cost of activity (COA), activity-thermoregulatory heat substitution, body mass, age, hematocrit, and a suite of morphological traits in female white-footed mice.

Trait	$T_a$ /method	Units	$n_{\text{obs}}$	$N_{\text{ID}}$	Mean	SD	Min	Max
Activity	10°C	$\text{m}\cdot\text{s}^{-1}$	117	46	0.054	0.035	0.003	0.170
Activity	22°C	$\text{m}\cdot\text{s}^{-1}$	117	46	0.029	0.024	0.003	0.106
DEE	10°C	$\text{kcal}\cdot\text{h}^{-1}$	117	46	0.667	0.328	0.207	1.489
DEE	22°C	$\text{kcal}\cdot\text{h}^{-1}$	117	46	0.473	0.218	0.152	1.093
COA	10°C	$\text{mLO}_2\cdot\text{m}^{-1}$	117	46	0.037	0.028	-0.016	0.124
COA	22°C	$\text{mLO}_2\cdot\text{m}^{-1}$	117	46	0.054	0.037	-0.013	0.164
Substitution	1 <sup>st</sup> method	$\text{mLO}_2\cdot\text{m}^{-1}$	117	46	0.017	0.017	-0.026	0.067
Substitution	2 <sup>nd</sup> method	$\text{mLO}_2\cdot\text{m}^{-1}$	117	46	0.014	0.018	-0.035	0.058
Body mass		g	117	46	19.51	3.45	13.80	33.50
Age		days	113	44	470.8	36.2	388.0	540.0
Hematocrit		ratio	104	44	0.459	0.087	0.250	0.682
Body length		cm	37	37	8.630	0.484	7.500	9.300
Tail length		cm	37	37	6.600	0.532	5.000	7.300
Surface area		$\text{cm}^2$	37	37	26.38	5.23	16.40	39.00
Tail mass		g	37	37	0.302	0.045	0.200	0.430
Fur mass		g	37	37	0.283	0.070	0.160	0.520
Skin mass		g	37	37	0.771	0.253	0.410	1.600
Heart mass		g	36	36	0.139	0.029	0.080	0.200
Liver mass		g	36	36	0.949	0.182	0.620	1.510
Kidneys mass		g	37	37	0.322	0.061	0.180	0.450
Right gastrocnemius mass		g	37	37	0.103	0.030	0.060	0.180

Descriptive statistics were calculated separately at the two temperatures ( $T_a$ ) for activity, DEE, and COA, separately for the two methods of estimation for substitution (see main text).

**Table 2.** Among-individual variance ( $V_I$ ), residual variance ( $V_e$ ), and consistency repeatability ( $R$ ) in body mass, daily energy expenditure (DEE), cost of activity (COA), and locomotor activity in 46 female white-footed mice.

Trait	Temperature	$V_I$	se	$\chi^2_{0:1}$	$P$	$V_e$	se	$R$	se
Body mass		0.913	0.203	106.8	<0.001	0.097	0.017	0.904	0.026
Activity	10°C	0.609	0.162	38.91	<0.001	0.331	0.059	0.648	0.087
Activity	22°C	0.422	0.133	20.29	<0.001	0.437	0.078	0.491	0.113
DEE	10°C	0.032	0.057	0.33	0.283	0.469	0.083	0.064	0.116
DEE	22°C	0.074	0.065	1.53	0.108	0.473	0.083	0.136	0.121
COA	10°C	0.087	0.085	1.14	0.143	0.609	0.108	0.126	0.126
COA	22°C	0.203	0.101	5.58	0.009	0.550	0.098	0.269	0.130

Estimates are from separate linear mixed models with metabolic cage and test sequence included as fixed effects. Variance and repeatability for activity, DEE, and COA are presented separately for observations taken at 22°C and 10°C. All traits are standardized to a mean of 0 and variance of 1. Significance of  $V_I$  was tested with a likelihood ratio test, which follows an equally weighted mixture of  $\chi^2$ -distributions with one and zero  $df$  ( $\chi^2_{0:1}$ ).

**Table 3.** Among-individual variance ( $V_I$ ), residual variance ( $V_e$ ), and repeatability ( $R$ ) in substitution in 46 female white-footed mice.

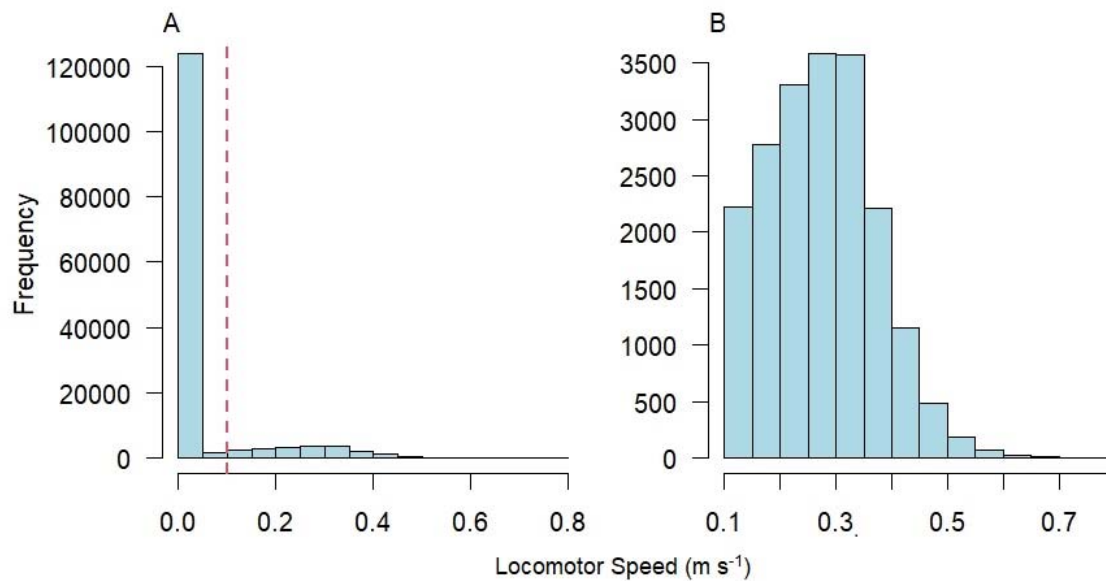
Calculation method	Covariates included?	Complete cases dataset?	$V_I$	se	$\chi^2_{0:1}$	$P$	$V_e$	se	$R$	se
A) first	no	no	0.306	0.134	7.66	0.003	0.672	0.119	0.313	0.131
B) second	no	no	0.279	0.133	6.18	0.006	0.724	0.128	0.278	0.132
C) first	yes	yes	0.074	0.112	0.51	0.237	0.529	0.119	0.123	0.184
D) First	no	yes	0.220	0.121	4.56	0.016	0.518	0.112	0.298	0.156

Each row provides the estimates from a different linear mixed model fitted to substitution values either calculated using A) the first method or B) the second method (described in text) as the dependant variable and only included nuisance variables (i.e., metabolic cage and test sequence) as fixed effects (consistency  $R$ ). In C), adjusted  $R$  was calculated from variance components estimated in a model that included the covariates listed in Table 4, but excluded individuals with a missing value for any of the covariate (i.e., a complete dataset without any missing values). In D), same as in A, but on the same complete cases dataset as in C. In all cases, substitution was standardized to a mean of 0 and variance of 1. Significance of  $V_I$  was tested with a likelihood ratio test, which follows an equally weighted mixture of  $\chi^2$ -distributions with one and zero  $df$  ( $\chi^2_{0:1}$ ).

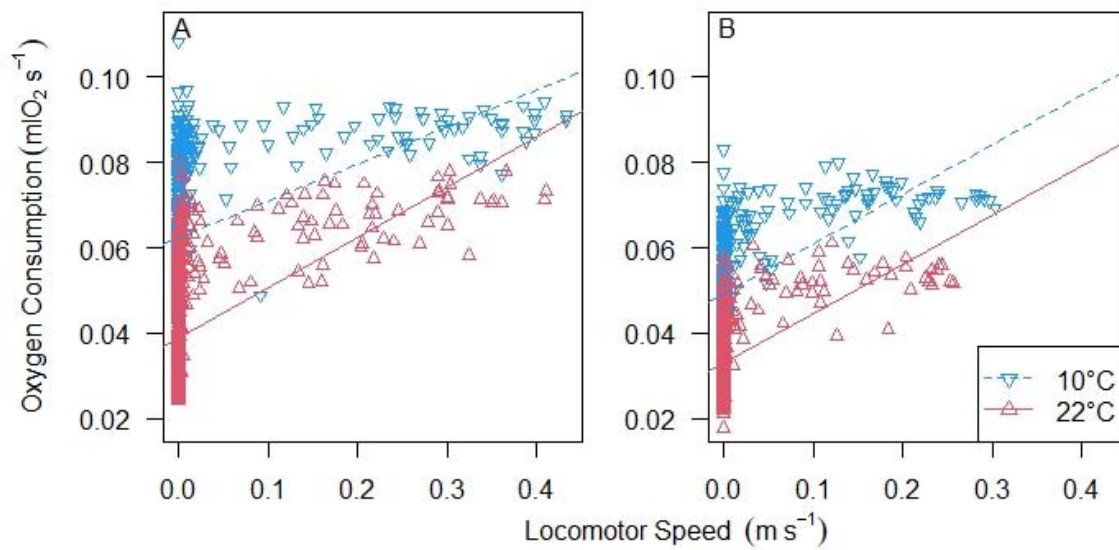
**Table 4.** Coefficient estimates, standard errors (SE),  $F$ -values, denominator degrees of freedom ( $df_{\text{den}}$ ), and  $P$ -values from a linear mixed model of heat substitution (dependent variable) that included a series of covariates included as independent variables.

Source	Estimate	SE	$F$	$df_{\text{den}}$	$P$
Intercept	0.270	0.348			
Body mass	0.223	0.144	2.38	31.4	0.1329
Age	0.293	0.195	2.27	23.5	0.1457
Torpor bouts	0.002	0.297	0.00	58.4	0.9940
Hematocrit	0.042	0.109	0.15	57.4	0.7000
<b>Body length</b>	<b>0.356</b>	<b>0.155</b>	<b>5.27</b>	<b>22.9</b>	<b>0.0312</b>
Tail length	-0.207	0.129	2.56	19.7	0.1256
<b>Surface area</b>	<b>-0.360</b>	<b>0.146</b>	<b>6.04</b>	<b>22.0</b>	<b>0.0223</b>
Tail mass	-0.260	0.192	1.83	30.3	0.1860
Fur mass	0.072	0.170	0.18	18.0	0.6753
Skin mass	0.020	0.135	0.02	21.1	0.8842
<b>Heart mass</b>	<b>0.333</b>	<b>0.156</b>	<b>4.53</b>	<b>25.8</b>	<b>0.0429</b>
Liver mass	-0.148	0.146	1.03	21.6	0.3207
Kidneys mass	-0.248	0.169	2.15	18.8	0.1596
Right gastrocnemius mass	0.044	0.113	0.15	18.3	0.7011
Resting metabolic rate	0.236	0.132	3.17	57.6	0.0804

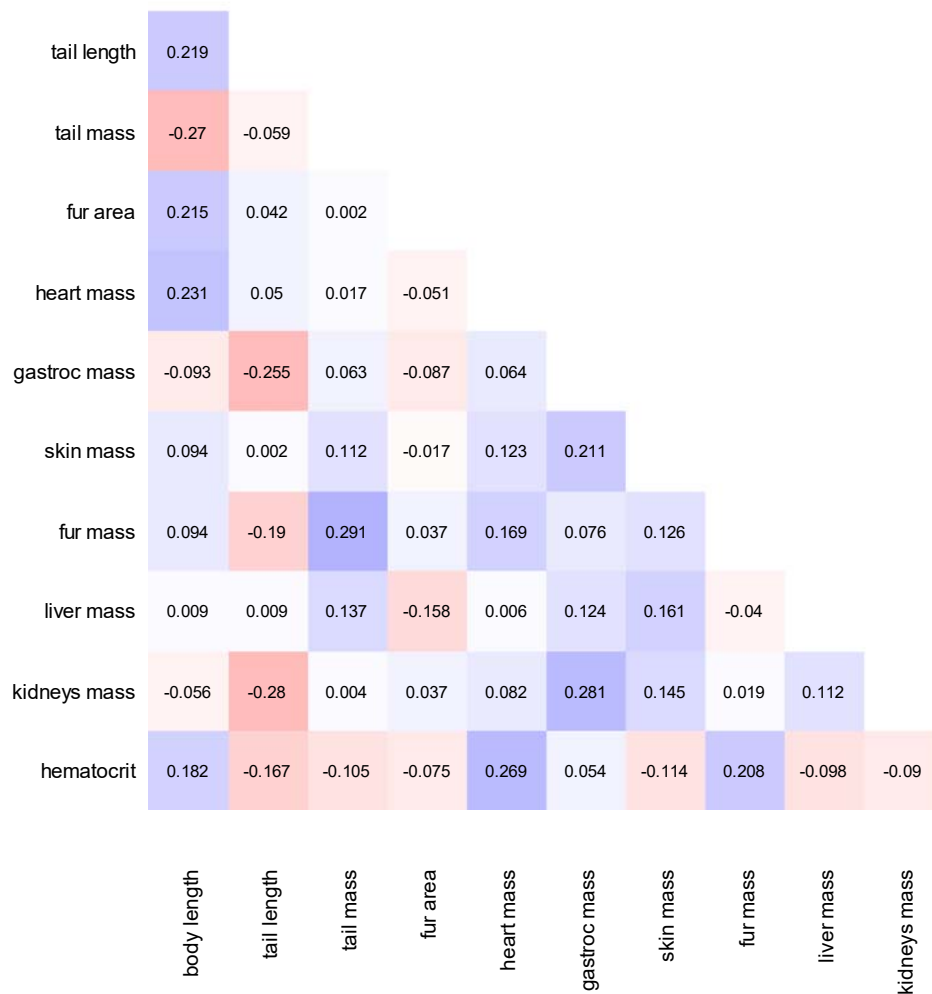
All variables were standardized to a mean of 0 and variance of 1. Fixed effects of test and cage not shown. Significant covariates ( $P < 0.05$ ) are bolded.



**Fig. S1.** Frequency distribution of locomotor speed ( $\text{m s}^{-1}$ ; as the sum of voluntary wheel-running and home-cage locomotion) observations for **A** the whole population across all tests and **B** speed observations greater than or equal to 0.1  $\text{m s}^{-1}$  (i.e., above the dashed line in A, the threshold above which a mouse was considered “active” in this study). Mice that had fewer than 5 observations above the threshold for each temperature condition for each test were excluded from analysis.



**Fig. S2.** Oxygen consumption ( $\text{ml s}^{-1}$ ) as a function of voluntary wheel-running and home-cage locomotor speed ( $\text{m s}^{-1}$ ) at  $22^{\circ}\text{C}$  (red upper triangles and solid line) and  $10^{\circ}\text{C}$  (blue lower triangles and dashed line) for **A** a single mouse exhibiting heat substitution (mouse #23517, test #2; see Table S1A) and **B** another mouse showing no substitution (mouse #23501, test # 1; see Table S1B).



**Fig. S3.** Correlation matrix of the various morphological and organ mass measurements included as covariates in the analysis of substitution.

**Table S1.** Parameter estimates from two representative multiple linear regression models used to quantify substitution (the “first method”), with metabolic rate (MR; oxygen consumption in mL per s) as a function of voluntary locomotor speed ( $\text{m s}^{-1}$ ), temperature (reference level 10°C), and their interaction for **A** single mouse exhibiting considerable heat substitution (mouse #23517, test #2) and **B** another mouse showing no substitution (mouse #23501, test # 1).

Source	A) mouse #23517, test #2		B) mouse #23501, test # 1	
	Estimate	± se	Estimate	± se
Intercept	0.0624	± 0.000529	0.0495	± 0.00035
Speed	0.0864	± 0.00602	0.116	± 0.00631
Temperature [22]	-0.0235	± 0.000749	-0.00165	± 0.000491
Speed × Temperature [22]	0.0312	± 0.0093	0.0000682	± 0.0104
The “speed × temperature” interaction term represent the difference between the slope of the MR-speed relationship at 22 °C vs 10°C, thus representing substitution. The “speed” term represents the slope (cost of activity) at 10°C. The slope at 22°C can be calculated by adding the “speed x temperature” interaction term to the “speed” term. See Figure S.2 for a visual representation of the regression lines fitted through these two sets of data.				

**Table S2.** Coefficient estimates, standard errors (SE),  $F$ -values, denominator degrees of freedom ( $df_{\text{den}}$ ), and  $P$ -values from a linear mixed model of heat substitution ( derived from second method, dependent variable) that included a series of covariates included as independent variables.

Source	Estimate	SE	$F$	$df_{\text{den}}$	$P$
Intercept	0.373	0.363			
Body mass	0.192	0.146	1.75	60.0	0.1915
Age	0.244	0.193	1.60	60.0	0.2106
Torpor bouts	0.001	0.321	0.00	60.0	0.9971
Hematocrit	-0.028	0.118	0.05	60.0	0.8166
<b>Body length</b>	<b>0.374</b>	<b>0.153</b>	<b>5.97</b>	<b>60.0</b>	<b>0.0175</b>
Tail length	-0.182	0.126	2.08	60.0	0.1547
<b>Surface area</b>	<b>-0.328</b>	<b>0.144</b>	<b>5.19</b>	<b>60.0</b>	<b>0.0263</b>
Tail mass	-0.333	0.194	2.96	60.0	0.0904
Fur mass	0.177	0.165	1.15	60.0	0.2870
Skin mass	-0.061	0.133	0.21	60.0	0.6489
<b>Heart mass</b>	<b>0.355</b>	<b>0.156</b>	<b>5.20</b>	<b>60.0</b>	<b>0.0262</b>
Liver mass	-0.201	0.144	1.95	60.0	0.1680
Kidneys mass	-0.230	0.165	1.96	60.0	0.1667
Right gastrocnemius mass	0.066	0.110	0.37	60.0	0.5479
Resting metabolic rate	0.151	0.144	1.10	60.0	0.2980

All variables were standardized to a mean of 0 and variance of 1. Fixed effects of test and cage not shown. Significant covariates ( $P < 0.05$ ) are bolded.