PERSPECTIVE

Making sense of heat tolerance estimates in ectotherms: lessons from Drosophila

Mauro Santos1, Luis E. Castañeda1,2 and Enrico L. Rezende*

1Departament de Genètica i de Microbiologia, Grup de Biologia Evolutiva (GBE), Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain; and 2Instituto de Ecología y Evolución, Facultad de Ciencias, Universidad Austral de Chile, Casilla 567, Valdivia, Chile

Summary

1. An increasing body of knowledge suggests that the estimation of critical upper thermal limits (CT\text{max}) is highly dependent on the experimental methodology employed. Here, we employ a theoretical approach to analyse how estimates of CT\text{max} (knock-down temperatures and times) are affected by measurement protocol.

2. Our model is able to reproduce the results of empirical studies on Drosophila melanogaster, suggesting that it adequately mimics organismal responses during assays. With simulated data sets, we also show that many experimental protocols result in unreliable and often highly biased estimates of CT\text{max} in Drosophila and possibly in other ectotherms.

3. The confounding effects of stochasticity, resource depletion (or fatigue) and short-term acclimatory responses are expected to be higher in longer assays, and therefore, short assays should be generally preferred. The experimental protocol of choice must also take into consideration the range in which measurement accuracy is not affected and the potential problems of thermal inertia in larger organisms.

4. Our findings justify previous concerns that the methodology may have a greater impact on estimates of CT\text{max} than the biological process under study, and explain why many studies on the subject have often reported inconsistent and even contradictory results.

Key-words: critical thermal limits, CT\text{max}, genetic correlation, Gompertz equation, repeatability, survival probability, thermotolerance

Introduction

Critical thermal limits are receiving increasing attention given their potential association with species distribution boundaries and responses to climate change (Deutsch et al. 2008; Pörtner & Farrell 2008; Angilletta 2009; Huey et al. 2009). Even though one might initially think that estimating the range of thermal tolerance with an acceptable degree of accuracy is an easy task, results in the literature indicate otherwise (Lutterschmidt & Hutchinson 1997). One of the primary problems pointed out in recent studies is that estimates of critical thermal limits can vary dramatically depending on the methodology employed (Elliott, Elliott & Allonby 1994; Elliott & Elliott 1995; Mora & Maya 2006; Terblanche et al. 2007; Chown et al. 2009; Mitchell & Hoffmann 2010). This is not entirely surprising because different protocols can result in varying levels of physiological stress, dehydration and so on, which may ultimately affect the estimation of thermotolerance limits (Rezende, Tejedo & Santos 2011). A more troubling aspect, however, is not the fact that estimates change with different methods, but the realization that patterns emerging from alternative protocols can be occasionally qualitatively different and, ultimately, contradictory.

In a recent study, Sgrò et al. (2010) uncovered contrasting latitudinal patterns for upper thermal limits (CT\text{max} and the proxy variables knock-down time and knock-down temperature) in Drosophila melanogaster populations from eastern Australia, which varied substantially with the experimental protocol used and resulted in latitudinal clines for thermal tolerance that were positive, negative and absent. They propose multiple adaptive explanations for these puzzling results and suggest that the underlying physiological and genetic mechanisms vary depending on the heat stress employed.

*Correspondence author. E-mail: enrico.rezende@uab.cat

© 2011 The Authors. Functional Ecology © 2011 British Ecological Society
The distinction between parameter and estimator used
Before we proceed, however, it is important to emphasize
representation of thermal tolerance in a broader sense at
whether any of these measurements provide an accurate
made. The fundamental question at this moment is
if sound extrapolations to natural populations are to be
performed by these researchers and their results emphasize
the other indices estimate? The impressive amount of work
from large-scale geographic patterns, which often include thermotolerance estimates obtained with
multiple experimental protocols in analyses (Addo-Bediako,
Chown & Gaston 2000; Chown 2001; Sunday, Bates & Dulvy
2011).

Given the discrepancies obtained by Sgrò et al. (2010),
and assuming that one of the measured variables is a good
proxy for heat tolerance, one is left to wonder: what do
the other indices estimate? The impressive amount of work
performed by these researchers and their results emphasize
the need to realize what is behind a ‘heat tolerance index’
if sound extrapolations to natural populations are to be
made. The fundamental question at this moment is
whether any of these measurements provide an accurate
representation of thermal tolerance in a broader sense at
all (which is what researchers presumably aim to estimate).
Before we proceed, however, it is important to emphasize
the distinction between parameter and estimator used
henceforth for the sake of clarity. The parameter \( CT_{\text{max}} \)
corresponds to the ‘real upper critical thermal limit’,
the maximum temperature that an organism might potentially
tolerate given its physiological condition in the absence of
any other hazard. Researchers attempt to quantify this
parameter employing two estimators, knock-down tempera-
tures and knock-down times. As discussed below, these
estimators may or may not accurately reflect \( CT_{\text{max}} \),
which can be very problematic when analysing the outcome
of experimental assays and documenting potential causal
effects.

What thermotolerance methods estimate

Common measures to quantify thermal tolerance include (i)
the static method, where organisms are placed acutely at a
constant stressful temperature \( T_{\text{stress}} \) and time to physical
incapacitation or knock-down time is recorded, and (ii) the
dynamic or ramping method, where temperature is increased
gradually at a rate \( \Delta T \) \(^\circ\)C \( \text{min}^{-1} \) from an initial tempera-
ture \( T_0 \) (usually not stressful) until individuals reach their
presumed critical thermal limit at time \( t = (CT_{\text{max}} - 
T_0)/\Delta T \) (min). Implicit in these protocols is the notion that
organisms are under thermal stress only above a given
threshold temperature \( T_{\text{threshold}} \), which explains why
static measurements subjecting D. melanogaster to 25 \(^\circ\)C
are interpreted as tests of desiccation or starvation resistance
(Hoffmann & Harshman 1999), whereas assays at 39 \(^\circ\)C
presumably quantify its tolerance to heat (Hoffmann,
Sørensen & Loeschcke 2003). In other words, thermal stress
is negligible at a body temperature \( T_b \) below \( T_{\text{threshold}} \) and
survival is conditioned by the depletion of resources such as
energy, water and metabolites, whereas at \( T_b \) above \( T_{\text{threshold}} \)
survival depends primarily on the capacity to withstand high
temperatures. Importantly, if assays last long enough to
have an impact on an organism’s physical condition, the
capacity to tolerate stressful temperatures will be partly
conditioned by resource depletion (Rezende, Tejedo & Santos
2011). This may be monitored and controlled to some extent
by measuring weight loss during assays.

Here, we employ a theoretical approach to analyse how
these factors interact, and compare our results against
published empirical data reported for D. melanogaster.
This is an expansion of our recent work (Rezende, Tejedo
& Santos 2011), which now incorporates the Gompertz
(1825) equation for mortality rates (a phenomenological
model widely used in demographic analyses; Finch 1990;
Hughes & Charlesworth 1994; Rauser et al. 2009). By
re-framing our model as a survival probability function,
we provide a general approach to assess how different experi-
mental conditions can affect estimates of thermal toler-
ance. This allows us (i) to visualize which factors may be
involved in an assay aimed to estimate thermotolerance
and (ii) to quantify and compare estimates of heat resis-
tance obtained with different methods under ‘controlled
conditions’. We introduce the formal mathematical for-
mulation underlying our model in the following sections and
provide the routines to run these analyses in MATLAB (V7,
MathWorks 2005; Appendix S1 in Supporting Informa-
tion) or R (http://www.r-project.org/; Appendix S2 in
Supporting Information).

The Gompertz equation as a tool in thermal
biology

Both the static and the dynamic methods involve estimating
at which point individuals collapse that, for simulation pur-
pose, can be assumed to be the lethal temperature. To obtain
the time \( t \) in which each individual succumbs to heat, we
modelled the probability of survival with the model proposed by
Gompertz (1825). In this model, mortality rates increase
exponentially with time:

\[
\mu(t) = Ae^{\alpha t},
\]

where \( \mu(t) \) is the time-specific mortality rate, \( A \) gives a base-
line mortality rate and \( \alpha \) is an age-dependent parameter.
The probability of any given individual surviving to time \( t \)
is as follows:

\[
P_t = \exp\left\{ \frac{A[1 - \exp(\alpha t)]}{\alpha} \right\},
\]

and the expected time each individual will survive is as fol-
loWS:

\[ E = \sum_{i=1}^{\infty} P_i. \]  

Eqn 3

To simulate the times in which an initial cohort of \( N_0 \) individuals would die, we now calculate the chance of dying \( u_i \) in the time interval \( t \rightarrow t + 1 \) as:

\[ u_i = 1 - \frac{P_{t+1}}{P_t}. \]  

Eqn 4

These demographic equations can be used to describe how survival probabilities change with the time elapsed since the beginning of the heat tolerance assays. Importantly, we must clarify that our model assumes the same cohort and, therefore, we are not dealing with changes in thermal tolerance as a function of age (e.g. Hollingsworth & Bowler 1966). Furthermore, so-called age effects associated with the age-dependent mortality-rate acceleration parameter \( \alpha \) are negligible (i.e. \( \alpha \) is close to zero) because assays are too short for individuals to ‘grow old’. As discussed below, temperature effects and individual differences in thermotolerance and metabolic rates, which will have an impact on the survival probability at any time interval during the assay, are incorporated in the baseline mortality-rate parameter \( A \) of the Gompertz equation (see Nusbaum, Mueller & Rose 1996).

**Mortality rates without thermal stress**

To illustrate how metabolic costs below \( T_{\text{threshold}} \) can be implemented in the Gompertz equation, we will make use of the data from Da Lage, Capy & David (1989) who recorded survival times during dehydration resistance assays in *D. melanoagaster* at nine temperatures ranging from 5 to 31 °C (only values equal to or above 14 °C will be considered to discard apparent deleterious effects of lower temperatures; see fig. 1 in Da Lage, Capy & David 1989). Empirical measurements support a significant impact of desiccation on heat tolerance (Maynard Smith 1957; Levins 1969; Parsons 1980; Block et al. 1994), and water depletion seems to be a more serious concern during thermal tolerance assays than food or energy depletion (table 1 in Rezende, Tejedo & Santos 2011; see also Huey et al. 1992, p. 492). Importantly, results described below are qualitatively robust, even though quantitative estimates will certainly change with the parameters employed to calibrate the model (e.g. other stocks may exhibit slightly different metabolic rates or desiccation resistance).

Temperature effects on the rate of resource expenditure can be quantified with the traditional \( Q_{10} \) factor:

\[ Q_{10} = \left( \frac{\text{MR}(T_2)}{\text{MR}(T_1)} \right)^{10/(T_2-T_1)}, \]  

Eqn 5

where temperatures \( (T_2 > T_1) \) are expressed in °C, \( \text{MR}(T_i) \) is the metabolic rate at the \( i \)th temperature, and \( Q_{10} \) is considered constant (recall we are, for the time being, within the species’ normal range of temperatures). Therefore, the amount of resources consumed at time \( t_i \) (the time interval here is 1 min) is calculated as follows:

\[ \text{EC}(t_i) = \sum_{i=1}^{t} MR_{t_i}(T). \]  

Eqn 6

which increases linearly with time when temperature \( T \) is kept constant as in the experiments by Da Lage, Capy & David (1989). The probability of survival of any individual to time \( t_i \) can now be derived from the Gompertz equation by assuming that its total water reserves at \( t_0 \) (Budget \( (t_0) \) decrease with time, with ‘intrinsic’ mortality rates \( A = \kappa / \text{Budget}(t_0) – \text{EC}(t_i) \)) (where \( \kappa \) is a constant and \( \kappa = \infty \) when Budget \( (t_0) – \text{EC}(t_i) \leq 0 \)):

\[ P(t_i) = \exp \left( \frac{\kappa}{\text{Budget}(t_0) – \text{EC}(t_i)} \times \frac{[1 – \exp(a t_i)]}{\alpha} \right). \]  

Eqn 7

We numerically computed the survival probabilities from eqn 7 in an initial cohort with \( N_0 = 24 \) flies as in the real experiments (Da Lage, Capy & David 1989) assuming an average weight of 1 mg and average metabolic rate of 4.2 mL O₂ g⁻¹ h⁻¹ at 18 °C (Berrigan & Partridge 1997). For each of the \( N_i \) survivors at any time, a random number between 0 and 1 was chosen from a uniform distribution. If the number was less than \( u_i \) (eqn 4), the individual died in min \( t + 1 \); otherwise, the individual lived. The process was repeated until all \( N_0 \) individuals had died.

Simulation results converge when both \( \alpha \) and \( \kappa \) tend to zero; hence for simplicity, here, we report the results obtained when both constants are set to \( 1 \times 10^{-6} \) (Fig. 1). We are able to numerically replicate the results from Da Lage, Capy & David (1989), suggesting that the simple Gompertzian mortality-rate function captures the essential features of their experiment by assuming that the total amount of water resources is limited and more or less constant across individuals. This model can now be employed as the baseline to quantify mortality rates in the absence of thermal stress, which allows us to estimate how stressful high temperatures affect survival.

**Incorporating thermal stress in the Gompertzian mortality-rate function**

In the absence of thermal stress, the expected survival time at 45 °C would correspond to 105.8 min (assuming \( Q_{10} = 2.5 \)) or 54.2 min (\( Q_{10} = 3.5 \)). Conversely, empirical measurements suggest that *D. melanoagaster* cannot tolerate more than 30 min at temperatures substantially lower such as 39 °C, and individuals die almost instantaneously at temperatures around 42 °C (Folk, Hoekstra & Gilchrist 2007; Sgrò et al. 2010). It is therefore clear that thermal stress, and not the depletion of resources, is the
cumulative water lost is calculated as a fraction of the total budget, which can be estimated from the desiccation assays in the absence of thermal stress (Fig. 1): for a water-deprived fly weighing 1 mg, with an average metabolic rate of 4.2 mL O₂ g⁻¹ h⁻¹ at 18 °C (Berrigan & Partridge 1997), this budget corresponds to 1356 μL O₂ (Q₁₀ = 2.5) or 171 μL O₂ (Q₁₀ = 3.5). As water reserves are depleted according to eqn 6 – notice that EC(t) increases linearly (static method) or exponentially (dynamic method) with time (see Rezende, Tejedo & Santos 2011) – the capacity to withstand high temperatures decreases and, in the absence of stochasticity, flies collapse at Tb = CT_max (t₁) (Appendix S3 in Supporting Information).

The probability of any given individual to survive at time t₁ described in eqn 7 must be modified to include thermal stress according to eqn 8 as:

$$P(t₁) = \begin{cases} \exp\left\{ \frac{\kappa}{[\text{Budget}_1(t₁) - \text{EC}(t₁)]} \times \frac{1 - \exp(\text{EC}(t₁))}{2} \right\} & \text{if } T_b < T_{\text{threshold}} \\ \exp\left\{ \frac{\kappa}{[\text{Budget}_1(t₁) - \text{EC}(t₁)]} \times \frac{1 - \exp(\text{EC}(t₁))}{2} \right\} & \text{if } T_b \geq T_{\text{threshold}} \end{cases}$$  

$$T_b = CT_{\text{max}}$$ can be easily modelled by noting that P₁ → 0 when A → ∞; hence, the effects of thermal stress can be incorporated in a modified function of mortality rates:

$$A = \begin{cases} \frac{\kappa}{[\text{Budget}_1(t₁) - \text{EC}(t₁)]} & \text{if } T_b < T_{\text{threshold}} \\ \frac{\kappa}{[\text{Budget}_1(t₁) - \text{EC}(t₁)]} \times \frac{1 - \exp(\text{EC}(t₁))}{2} & \text{if } T_b \geq T_{\text{threshold}} \end{cases}$$  

with \( \frac{[CT_{\text{max}1}(t₁) - T_b]}{[CT_{\text{max}1}(t₁) + T_b]} = 0 \) when \( CT_{\text{max}1}(t₁) - T_b \leq 0 \). When \( T_b \geq T_{\text{threshold}} \), the numerator of eqn 8 still assumes a temperature dependence of metabolic rates, but the exponentially decreasing function in the denominator incorporates the additional effect of thermal stress (see also supporting information in Rezende, Tejedo & Santos 2011).

Cumulative water lost is calculated as a fraction of the total budget, which can be estimated from the desiccation assays in the absence of thermal stress (Fig. 1): for a water-deprived fly weighing 1 mg, with an average metabolic rate of 4.2 mL O₂ g⁻¹ h⁻¹ at 18 °C (Berrigan & Partridge 1997), this budget corresponds to 1356 μL O₂ (Q₁₀ = 2.5) or 171 μL O₂ (Q₁₀ = 3.5). As water reserves are depleted according to eqn 6 – notice that EC(t₁) increases linearly (static method) or exponentially (dynamic method) with time (see Rezende, Tejedo & Santos 2011) – the capacity to withstand high temperatures decreases and, in the absence of stochasticity, flies collapse at Tb = CT_max (t₁) (Appendix S3 in Supporting Information).

The probability of any given individual to survive at time t₁ described in eqn 7 must be modified to include thermal stress according to eqn 8 as:

$$P(t₁) = \begin{cases} \exp\left\{ \frac{\kappa}{[\text{Budget}_1(t₁) - \text{EC}(t₁)]} \times \frac{1 - \exp(\text{EC}(t₁))}{2} \right\} & \text{if } T_b < T_{\text{threshold}} \\ \exp\left\{ \frac{\kappa}{[\text{Budget}_1(t₁) - \text{EC}(t₁)]} \times \frac{1 - \exp(\text{EC}(t₁))}{2} \right\} & \text{if } T_b \geq T_{\text{threshold}} \end{cases}$$  

This equation provides a formal description of how survival probabilities change during the course of the assay as a function of both time t₁ (i.e. individuals are losing water during the trial and they should eventually desiccate to death in the absence of thermal stress) and body temperature T_b (i.e. the effects of thermal stress may eventually result in death regardless of the total amount of water that remains available). The fundamental point to be stressed is that the time t₁ and body temperature T_b at which an individual collapses involve both a deterministic component (heat stress and physical condition) and a stochastic component (eqn 4), which are described with a survival probability curve that decreases from 1 to 0 (eqn 10).

Armed with this mathematical formulation, we now analyse whether the model is able to replicate results from different studies of thermostolerance in D. melanogaster (the data were often retrieved from the original plots with GRAPHCLICK.
Modelling survival probability in heat resistance assays

3.0; http://www.arizona-software.ch/graphclick/). Our aim is to show that the Gompertzian mortality-rate function provides a reasonable fit to empirical data obtained with different protocols and can be employed to elucidate which factors may underlie discrepancies across measurements.

Theoretical predictions and empirical data

The impact of thermal stress can be simulated and contrasted against the scenario in which this effect is negligible using eqn 10 instead of eqn 7 (Fig. 2). With \( N_0 = 24 \) as above and \( T_b \geq T_{\text{threshold}} = 36 \, ^\circ \text{C} \), average survival time drops from 136.5 min at 36 \(^\circ\)C to instantaneous death at \( CT_{\text{max}} = 41 \, ^\circ \text{C} \) as expected, being 26.6 min at 39 \(^\circ\)C and reasonably close to empirical results reported for static assays at this temperature (Parkash, Sharma & Kalra 2010; Sgro et al. 2010). Importantly, the sudden drop at 36 \(^\circ\)C (Fig. 2a) reflects the conditional nature of eqn 10, and thermal effects are probably more gradual in reality. Nonetheless, these analyses illustrate how \( T_{\text{threshold}} \) can be estimated empirically, by measuring survival times in flies exposed at different temperatures and determining the region of the curve in which survival drops below levels expected from \( Q_{10} \) effects.

Can our model replicate the results reported by other studies? Comparisons between thermotolerance obtained from several empirical studies in \( D. \) melanogaster and the values obtained from eqn 10 show that the proposed theoretical model can reproduce both the qualitative patterns – e.g. the trend that knock-down temperature generally increases with increasing \( T_b \) and \( \Delta T \) in ramping assays (see Terblanche et al. 2007) – and numerical results quite well (Fig. 3). A remarkable exception occurs with the exceedingly high values reported by Sgro et al. (2010) with their fast ramping protocol \( T_b = 25 \, ^\circ \text{C} \) and \( \Delta T = 0.1 \, ^\circ \text{C} \text{min}^{-1} \) that, interestingly, is also at odds with results from their static assay (this is discussed in detail below). These results suggest that in spite of the potential differences in upper thermal limits and other physiological traits among measured stocks, our model can successfully replicate the overall pattern obtained from empirical studies.

Correlation between \( CT_{\text{max}} \) and heat tolerance estimates

We now explore how the stochasticity inherent to thermotolerance assays may affect the outcome of experiments, employing simulated data sets. Let us consider for a moment what a researcher implicitly assumes when concluding that \( CT_{\text{max}} \) corresponds to the measured knock-down temperature (or any similar operational definition): namely, that the probability of surviving at \( T_b < CT_{\text{max}} \) is close to one and suddenly becomes zero when \( T_b = CT_{\text{max}} \) which is utterly unrealistic. A more reasonable scenario is that the probability of surviving decreases gradually as \( T_b \) approaches \( CT_{\text{max}} \) as shown in Fig. 2b (and described empirically in the leaf-cutter ant \( Atta sexdens rubropilosa \); fig. 3a in Angilletta et al. 2007). In other words, genetically identical individuals that can tolerate in the very best of cases a temperature of 41 \(^\circ\)C will likely collapse at \( T_b \) ranging from, say, 38.5 to 41 \(^\circ\)C because of stochastic circumstances that researchers cannot control. For instance, a recent experiment aimed to detect genetic differences for knock-down temperature among various karyotypes in \( D. \) subobscura used 18 independent isogenic lines, where the only source of phenotypic variation between individuals within a given line is...
among isogenic lines was 0.183, which corresponds to the ‘repeatability’ of the estimate (Sokal & Rohlf 1995; Falconer & Mackay 1996). The remaining variation possibly involves stochasticity and measurement error. From a statistical perspective, stochasticity will always downward bias heat tolerance estimates because CTmax is by definition the physiological limit (upper boundary).

We can now ask to what extent the values of knock-down temperatures (or knock-down times) reflect the parameter CTmax that researchers would like to estimate. This can be readily carried out with our model, by setting CTmax values a priori and then determining knock-down temperatures (or times) obtained in our computer-generated heat tolerance assay (note that, assuming no variation in metabolic rates and total reserves between individuals and in the absence of stochasticity, the correlation between these values should be one; Appendix S3 in Supporting Information). This approach suggests that a substantial fraction of the variation in CTmax is lost owing to stochasticity (Table 1).

Correlation coefficients in Table 1 describe the amount of signal that is lost in each assay; hence, the expected correlation between two independent assays 1 and 2 corresponds to the product between correlations r1 × r2. This follows from a straightforward application of the theory of path coefficients (Li 1976) because:

\[ r(ht_1, ht_2) = r(CTmax, ht_1) \times r(CTmax, ht_2), \]

where ht1 and ht2 are the estimated heat tolerance (knock-down temperatures or knock-down times) in assays 1 and 2 and CTmax is the common causal variable. Numerical results from simulations employing the same initial cohort are given in Table 2 and represent an upper bound on the accuracy of these methods to appropriately estimate CTmax, its repeatability and heritability and on the correlations between different estimators. These results are conservative because they ignore measurement error, thermal inertia, hardening or individual variation in metabolism, which would contribute to overall ‘measurement noise’.

Unfortunately, we cannot know for certain the real magnitude of stochasticity effects because they depend on the shape and parameters of the survival probability function (Fig. 2b), yet the repeatability of 0.183 among isogenic lines of D. subobscura (a similar value was obtained in D. buzzatii; Krebs & Loeschcke 1997) emphasizes that the effects of experimental noise can be substantial. Nonetheless, some qualitative conclusions can be generalized because all realistic functions should ultimately describe a decrease in survival probability and converge to zero as Tb approaches CTmax. If measurement error and effects of thermal inertia and short-term plasticity do not differ between protocols (which are important caveats to keep in mind during experimental design), fast assays are expected to provide a more accurate representation of the underlying genetic variation in heat tolerance than slow assays (Table 1) for two reasons. First, metabolic costs are increasingly higher in long assays (see Rezende,
Modelling survival probability in heat resistance assays

Table 1. Pearson’s correlation between CT\text{max} and estimated heat knock-down temperature (ramping protocols) or heat knock-down time (static assays)

<table>
<thead>
<tr>
<th>Experimental settings</th>
<th>Source</th>
<th>Correlation (r ± SE)*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ramping</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T_0 = 30 \degree C; \Delta T = 0.24 \degree C \text{ min}^{-1}$</td>
<td>Folk, Hoekstra &amp; Gilchrist (2007)</td>
<td>$0.5939 \pm 0.0114$</td>
</tr>
<tr>
<td>$T_0 = 20 \degree C; \Delta T = 0.1 \degree C \text{ min}^{-1}$</td>
<td>Chown et al. (2009)</td>
<td>$0.5188 \pm 0.0121$</td>
</tr>
<tr>
<td>$T_0 = 20 \degree C; \Delta T = 0.25 \degree C \text{ min}^{-1}$</td>
<td>Chown et al. (2009)</td>
<td>$0.5343 \pm 0.0120$</td>
</tr>
<tr>
<td>$T_0 = 20 \degree C; \Delta T = 0.5 \degree C \text{ min}^{-1}$</td>
<td>Chown et al. (2009)</td>
<td>$0.6022 \pm 0.0113$</td>
</tr>
<tr>
<td>$T_0 = 28 \degree C; \Delta T = 0.06 \degree C \text{ min}^{-1}$</td>
<td>Mitchell &amp; Hoffmann (2010)</td>
<td>$0.5312 \pm 0.0120$</td>
</tr>
<tr>
<td>$T_0 = 25 \degree C; \Delta T = 0.1 \degree C \text{ min}^{-1}$</td>
<td>Sgrò et al. (2010)</td>
<td>$0.5319 \pm 0.0120$</td>
</tr>
<tr>
<td><strong>Static</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T_{\text{stress}} = 38 \degree C$</td>
<td>Mitchell &amp; Hoffmann (2010); Parkash, Sharma &amp; Kalra (2010)</td>
<td>$0.5781 \pm 0.0115$</td>
</tr>
<tr>
<td>$T_{\text{stress}} = 39 \degree C$</td>
<td>Parkash, Sharma &amp; Kalra (2010); Sgrò et al. (2010)</td>
<td>$0.7070 \pm 0.0100$</td>
</tr>
</tbody>
</table>

Heat tolerances estimated from simulated cohorts with $N_0 = 5000$ flies. Variation for CT\text{max} was sampled from a normal distribution with mean $\mu = 41 \degree C$ and standard deviation $\sigma = 0.5 \degree C$ or $\sigma = 1 \degree C$. Metabolic rate was kept constant across individuals at 4.2 mL O$_2$ g$^{-1}$ h$^{-1}$ at 18 $\degree C$ (Berrigan & Partridge 1997). Survival probabilities derived form eqn (10) with $T_{\text{threshold}} = 36 \degree C$ and parameters $\kappa = \alpha = 1 \times 10^{-5}$. The best $Q_{10} = 3.5$ fit to Da Lage, Capy & David (1989) empirical values (Fig. 1) was used.

*Standard error computed as $\sqrt{1-r^2}/(n-2)$ (Sokal & Rohlf 1995).

Table 2. Pearson’s correlation of estimated heat tolerance within assays (repeatability) and between assays using static and ramping protocols

<table>
<thead>
<tr>
<th>Experimental settings</th>
<th>Correlation (r ± SE)*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trial 1</strong></td>
<td><strong>Trial 2</strong></td>
</tr>
<tr>
<td>$T_{\text{stress}} = 38 \degree C$</td>
<td>$T_{\text{stress}} = 38 \degree C$</td>
</tr>
<tr>
<td>$T_{\text{stress}} = 39 \degree C$</td>
<td>$T_{\text{stress}} = 39 \degree C$</td>
</tr>
<tr>
<td>$T_0 = 20 \degree C$ and $\Delta T = 0.1 \degree C \text{ min}^{-1}$</td>
<td>$T_0 = 20 \degree C$ and $\Delta T = 0.1 \degree C \text{ min}^{-1}$</td>
</tr>
<tr>
<td>$T_0 = 20 \degree C$ and $\Delta T = 0.5 \degree C \text{ min}^{-1}$</td>
<td>$T_0 = 20 \degree C$ and $\Delta T = 0.5 \degree C \text{ min}^{-1}$</td>
</tr>
</tbody>
</table>

Heat tolerances estimated from simulated cohorts with $N_0 = 5000$ flies. Variation for CT\text{max} was sampled from a normal distribution with mean $\mu = 41 \degree C$ and standard deviation $\sigma = 0.5 \degree C$ or $\sigma = 1 \degree C$. Metabolic rate was kept constant across individuals at 4.2 mL O$_2$ g$^{-1}$ h$^{-1}$ at 18 $\degree C$ (Berrigan & Partridge 1997). Survival probabilities derived form eqn (10) with $T_{\text{threshold}} = 36 \degree C$ and parameters $\kappa = \alpha = 1 \times 10^{-5}$. The best $Q_{10} = 3.5$ fit to Da Lage, Capy & David (1989) empirical values (Fig. 1) was used.

*Standard error computed as $\sqrt{1-r^2}/(n-2)$ (Sokal & Rohlf 1995).

of temperature can be used to approximate CT_max, which appears to
be seriously underestimated in ramping protocols with slow heating
rates (see also Cooper, Williams & Angilletta 2008).

Fig. 4. Simulated distributions of thermal sensitivities of knock-
down temperatures in dynamic assays (bars) and average knock-
down times in static assays (black points) obtained from cohorts with
initially N_0 = 500 flies each, assuming variation for both CT_max
(mean ± SD: 41 ± 0.5°C) and metabolic rate (4.2 ± 0.4 mL O_2 g^{-1} h^{-1}
at 18°C). Survival times were derived from eqn (10) with parameters
\( \kappa = \alpha = 1 \times 10^{-4} \) \( (Q_{10} = 3.5) \).
When temperature increases from \( T_0 = 28 \) °C at ramping rates
\( \Delta T = 0.1, 0.5 \) or \( 10 \) °C min^{-1}, average knock-down temperature
increases from 37.7 to 38.6 and 38.4 °C, respectively. The plot illus-
trates that knock-down times in static assays over an increasing range
of temperature can be used to approximate CT_max, which appears to
be seriously underestimated in ramping protocols with slow heating
rates (see also Cooper, Williams & Angilletta 2008).

for populations/species with high CT_max than for their coun-
terparts with low CT_max, resulting in empirical clines being
consistently underestimated in comparison with the original
‘genetically determined’ cline in thermal tolerance (Fig. 5). Conse-
quently, several empirical latitudinal gradients of thermotolerance
might have underestimated the magnitude of the actual biological pattern, especially those involving long
assays in which the cumulative effects of thermal stress and
resource depletion may have affected survival probabilities
(e.g. subjecting small drosophilids to a static thermal stress
lasting 24 h; Kimura 2004).

In summary, the different indices proposed to estimate thermotolerance inherently involve a substantial amount of noise that, alas, may not be easily removed. We suspect that
this problem is pervasive and underlies many of the low phe-
notypic correlations between estimates of thermotolerance
(Berrigan & Hoffmann 1998; Sgro et al. 2010) and the absence of correlated responses in selection experiments (Gil-
christ, Huey & Partridge 1997; Hoffmann et al. 1997). It may
also explain why measurements of thermal tolerance are
seemingly independent within species but provide consistent
results across species (e.g. Berrigan & Hoffmann 1998; Berri-
gan 2000; Mitchell & Hoffmann 2010), because statistical
power increases with the variability in CT_max (the stronger
the signal, the more likely it is to detect it). For instance, the
variance in mean knock-down times with a static assay at
38 °C (\( \sigma^2 \approx 141 \) min^{-2}; fig. 3a in Mitchell & Hoffmann 2010)
across Drosophila species corresponds to roughly 57 times the
variance across population means for D. melanogaster mea-
sured at 39 °C with a similar protocol (\( \sigma^2 \approx 2.5 \) min^{-2}; fig. 1a
in Sgro et al. 2010). For knock-down times with comparable
ramping assays, with heating rates of 0.06 °C min^{-1}, this ratio
increases to over 90 times (\( \sigma^2 \approx 867 \) min^{-2} vs. 9.4 min^{-2}).
Therefore, discrepancies between inter- and intraspecific

Fig. 5. Simulated latitudinal clines in knock-down temperatures measured with dynamic assays employing two ramping conditions. We assumed
that the average population CT_max decreases with latitude at a rate of 0.1 °C per degree of latitude, a variance in CT_max of \( \sigma^2 = 0.7 \) between pop-
ulation means and \( \sigma^2 = 0.3 \) within populations (resulting in total variance \( \sigma^2 = 1.0 \)). From this original cline, we sampled 30 individuals per
population, calculated their mean knock-down temperatures during dynamic assays (\( T_0 = 20 \) °C) and the slope of the resulting cline. This proce-
dure was repeated 100 times. The error bars represent the 95% confidence intervals for population means, and the histograms depict the distribu-
tion of the slopes obtained in these ‘independent latitudinal studies’. The plot shows that the effects of stochasticity are more pronounced in the
slow ramping assay, having an impact on both the intercept and the slope of the latitudinal cline. We assumed a \( Q_{10} = 3.5 \) and a metabolic rate
of 4.2 mL O_2 g^{-1} h^{-1} at 18°C.
results may reflect a problem of statistical power resulting from contrasting signal/noise ratios (Table 1), and the absence of correlation between estimates does not necessarily support independent genetic and physiological pathways.

Physiological regulation and phenotypic plasticity

Overall, our model explains why different estimates of thermotolerance are not expected to be strongly correlated even when they share the same genetic basis, but additional assumptions are necessary to reproduce the negative correlations across estimates reported by Sgrò et al. (2010) (and the change in sign of the latitudinal clines depending on the estimate employed). One possible explanation for their patterns would involve a subjacent cline in metabolic rates, in which a positive relationship between knock-down temperatures estimated with an acute protocol (where assays last 5 min) becomes negative in longer assays because of the differential impact of resource depletion in thermal tolerance (eqn 9). This corresponds to the ‘thermal compensation hypothesis’ (Clarke 1993; Irlisch et al. 2009), which posits that cold-adapted organisms have higher metabolic rates than their warm-adapted counterparts, and finds some empirical support in D. melanogaster flies collected along the Australian cline (Berrigan & Partridge 1997). Another possibility is that flies acclimate during experiments (plastic responses have been ignored in our model but could be implemented with a function describing an increase in \( T_{\text{threshold}} \) or \( CT_{\text{max}}(t) \) with time; see Rezende, Tejedo & Santos 2011) and that populations with low thermotolerance during an acute assay show a disproportionately higher acclimatory response during longer measurements. Accordingly, the threshold temperature to induce hardening is lower in cold- vs. warm-adapted lines of Drosophila (Cavicchi et al. 1995), and cold-adapted populations of D. melanogaster – as well as porcelain crab species (Stillman 2003) – exhibit higher acclimatory responses than their warm-adapted counterparts when differences between acclimation temperatures are held constant (Parkash, Sharma & Kalra 2010). In addition, Sgrò et al.’s (2010) results also provide some support for both possibilities.

We suspect that their populations from high latitudes were thermally challenged at the common garden temperature of 25 °C, hence recruiting more molecular chaperones under basal conditions – and consuming more ATP to maintain proteins folded (Hochachka & Somero 2002) – than their low-latitude counterparts. This would help to explain why high-latitude populations were able to withstand higher temperatures during an acute thermal challenge (their fig. 1F) but not during a static assay at 39 °C (their fig. 1A) and why they exhibited a decreased hardening response to heat shock (their fig. 1A–C). Differences in hardening responses across populations can be seen in their ramping trials, which ultimately capture the change in sign of the cline in real time: in the fast ramping assay, populations appear to be acclimating and the cline is flat (their fig. 1D), whereas in the slow ramping assay, they seem to be fully acclimated and the cline eventually becomes negative and congruent with the static assay (their fig. 1E). If these conjectures are correct, their populations from low latitudes should have a higher \( CT_{\text{max}} \) than those from high latitudes everything else being equal, which would be an adaptive pattern lying beneath their contrasting clines.

Other observations support this hypothesis. First and most importantly, one should expect differences in heat tolerance between populations to be more pronounced when stocks are raised at lower temperatures, which is precisely what Parkash, Sharma & Kalra (2010) reported. These authors measured knock-down times in D. melanogaster with the same static protocol employed by Sgrò et al. (2010) and observed substantially steeper altitudinal clines when stocks were maintained at low temperatures (about 2-3 times steeper at 17 °C than at 25 °C; fig. 1E in Parkash, Sharma & Kalra 2010). This result provides unambiguous evidence that their ‘cold-adapted’ populations are thermally challenged and relied more heavily on acclimatory responses to survive at 25 °C, and we postulate that the same might be true for the populations studied by Sgrò et al. (2010). Second, knock-down times calculated from knock-down temperatures reported during the fast ramping assay (fig. 1D in Sgrò et al. 2010) suggest that many southern populations withstood temperatures above 39 °C for more than 20 min (28 min in one case, with flies collapsing at approximately 41.8 °C). Some of these estimates are up to 6 or 7 min higher than reported basal knock-down times at 39 °C (their fig. 1A) and can only be explained by hardening. Thus, hardening was apparent in both fast and slow ramping assays, contrary to the suggestion of Sgrò et al. (2010) that ‘the faster rate of temperature increase may occur too quickly for any hardening response to occur’ (p. 2491; if anything, these results illustrate the perils of employing two different units, time and temperature, to quantify heat tolerance). Finally, many genes encoding heat-shock proteins seem to be overexpressed in cold- vs. warm-adapted thermal lines of D. subobscura measured at the optimum temperature of 18 °C (additional file 1 in Laayouni et al. 2007). Chaperones might thus respond to acclimation conditions, yet the generality of this pattern remains to be studied.

In summary, thermotolerance seems to be very sensitive to environmental conditions. Patterns reported by Sgrò et al. (2010) suggest that physiological accommodation to maintenance temperatures and during experimental assays can dramatically affect the outcome of very carefully designed studies, and provide an additional source of noise that may impair comparisons across methods. Because proteins exist in a gradient of conformational states at any given temperature (i.e. a gradient of denaturation, see p. 310 in Hochachka & Somero 2002), we speculate that ectothermic species modulate the concentration of chaperones at varying temperatures just as endotherms regulate metabolic rates to maintain body temperature constant (this response is probably not nearly instantaneous as in endotherms). This would explain the patterns reported by these studies (Parkash, Sharma & Kalra 2010; Sgrò et al. 2010) and also suggest that heat-shock responses represent a specific case of a more general
mechanism of physiological regulation to maintain homoeostasis at different temperatures.

Obtaining more reliable estimates of CT\textsubscript{max}

We have employed Drosophila because the physiological parameters required for quantitative analyses, as well as the number and scope of empirical studies (latitudinal clines, selection experiments, heritability and repeatability assessment, etc.), are readily available for this study model. However, the proposed conceptual framework is potentially relevant for other ectothermic species. We discussed four potential factors—other than measurement error per se—that may affect the outcome of tolerance assays and subsequent conclusions, which are absolutely general: (i) stochasticity, (ii) depletion of resources, (iii) hardening during assays and (iv) acclimation to maintenance conditions. Notably, only stochasticity can be considered a true source of error because it results in a biased estimator of CT\textsubscript{max}. The other three factors involve real biological changes in CT\textsubscript{max} (i.e. on the parameter itself), which is expected to decrease when organisms are in poor physical conditions or increase during hardening (see Rezende, Tejedo & Santos 2011). They are only considered a source of noise because many studies have failed to account for their potentially confounding effects, which will vary in magnitude and relative importance depending on the experimental conditions and the study model (e.g. large organisms are expected to have a lower mass-specific metabolic rate and increased reserves; hence, decaying physical conditions in a 2-h measurement is probably less of a problem for a larger insect than a Drosophila). Although it might be virtually impossible to entirely control for these factors, our results provide a guideline on how to minimize their effects and obtain more reliable thermotolerance data (see also Lutterschmidt & Hutchinson 1997).

The experimental protocol should be as brief as possible to minimize stochasticity, resource depletion and hardening. This choice must, of course, take into consideration the range in which measurement accuracy is not affected and the potential problems associated with thermal inertia between ambient and body temperature in larger organisms (Lutterschmidt & Hutchinson 1997). Although this recommendation seems to ignore ecological realism (see Chown et al. 2009), simulations show that longer and ‘ecologically realistic’ assays can result in very unreliable estimates of CT\textsubscript{max} owing to increased levels of noise associated with the factors discussed above (see also Rezende, Tejedo & Santos 2011). These observations provide an objective criterion to assess which protocols provide more reliable estimates of CT\textsubscript{max}, which outweighs the subjectivity inherent to ‘ecological realism’ as a criterion to design experiments of thermotolerance.

With regard to common garden conditions, cold temperatures should be preferred over warm temperatures to minimize confounding acclimatory effects. Comparative biologists have pointed out the limitations of the common garden approach when comparing different species (Garland 2001; Garland, Bennett & Rezende 2005), and a similar problem may obscure comparisons between populations genetically adapted to local conditions. Parkash, Sharma & Kalra (2010) reported more pronounced population differences in knock-down time across stocks maintained at colder temperatures that, we postulate, reflect more faithfully basal genetic differences in CT\textsubscript{max}. This raises the issue of ‘how cold is cold enough’, which does not have a straightforward answer because the impact of maintenance temperature is both species- and context-dependent (one obvious problem is that colder temperatures may also eventually impinge on deleterious effects; e.g. see fig. 3 in Rezende, Tejedo & Santos 2011). As a provisional rule of thumb, we recommend selecting temperatures at the lower end of the thermal range in the latitudinal cline under study.

Researchers should also recall that thermotolerance varies as a function of temperature and time. Even though this fact is evident in static assays (an organism may be able to tolerate 38°C for 20 min and 40°C for 5 min), it is not nearly as obvious in ramping experiments with varying T\textsubscript{m} and heating rates (see also Terblanche et al. 2007). Working with estimates in different units (of time or temperature) can be cumbersome and obscure analyses, hindering comparative efforts and the quest for general patterns. Thus, we recommend transforming all estimates into a common currency during analyses (otherwise, acclimatory responses and other relevant patterns may be missed, as illustrated above) and reporting both knock-out times and temperatures for transparency.

Concluding remarks

Several researchers have raised concerns regarding the impact of methodology on estimates of CT\textsubscript{max} (Lutterschmidt & Hutchinson 1997; Terblanche et al. 2007; Chown et al. 2009), and our theoretical approach justifies these concerns. There is a clear distinction between the parameter that researchers wish to measure (CT\textsubscript{max}) and the estimator they actually measure (knock-down temperatures and knock-down times). Our analyses show that estimates of CT\textsubscript{max} can be unreliable and often highly biased, which might affect a substantial fraction of the empirical data published to date. This problem should not be taken lightly: it impacts research across different branches of thermal biology, from studies on the genetic and physiological basis of thermal tolerance (heat-shock responses, heritability estimates and correlated evolutionary responses) to analyses of broad-scale geographic trends.

We have discussed how comparisons between estimates within a single study are not to be trusted, let alone comparisons across studies. Whereas one comparative study including 30 species of drosophilids across a gradient spanning nearly 20° of latitude reported lethal temperatures (defined as the temperature at which 50% of the population survived) ranging between 31 and 35°C after 24 h of exposure (Kimura 2004), another study in the tsetse fly Glossina palpalis reported lethal temperatures of 37.9 and 35.6°C following 1 and 3 h of exposure, respectively (Terblanche et al. 2008). In essence, a 2-h difference in exposure time in one study results in an effect equivalent to roughly 50% of the maximum
difference between species reported in another study (why both the mean and the genetic variance in CT_{max} are expected to decrease in long assays is also discussed in Rezende, Tejedo & Santos 2011). These estimates are clearly not comparable, yet they have been included in the most recent global analysis of thermal tolerance (Sunday, Bates & Dulvy 2011). Importantly, controlling for experimental differences may not be easily accomplished with standard statistical procedures (e.g. adding duration of assays, ramping rates or rearing temperature as covariables) because effects of different sources of error may be nonlinear (Fig. 3), they likely differ across populations or species, and interactions can be complex and give rise to opposing trends (c.f. fig. 2a in Terblanche et al. 2007 and fig. 3a in Nyamukondiwa & Terblanche 2010).

At this stage, it seems strikingly obvious that the method used may have a greater influence on [estimates of] CT_{max} than real species differences’ (p. 1569 in Lutterschmidt & Hutchison 1997; our addition between brackets) and this is also true for acclimatory effects (Chown et al. 2009). The unfortunate conclusion is that many of the patterns reported in the literature, as well as the inferences based on these patterns, should be revisited in the light of results described here. This does not imply that researchers should immediately ‘dismiss past generalities concerning interspecific and acclimation-related variation in critical thermal limits’ (p. 138 in Chown et al. 2009). Nonetheless, the empirical and theoretical evidence is compelling and alarming enough to put these past generalities under scrutiny.

Acknowledgements

We thank Ary Hoffmann and Carla Sgro for helpful discussions and Jennifer Sunday for kindly providing a copy of her manuscript in advance of publication. We are also very grateful to J. Malcolm Elliott for bringing to our attention some of the ‘old’ literature on the effects of different heating rates in fish species and how to assess their upper critical thermal limits and precision of estimates. The manuscript greatly benefited from the insightful comments raised by three anonymous reviewers and the editors. M.S. is supported by grant CGL2010-15395 from the Ministerio de Ciencia e Innovación (Spain) and by the ICREA Acadèmia program. E.L.R. is supported by a Ramón y Cajal contract and by grant BFU2009-07564 from the Ministerio de Ciencia e Innovación. Financial support by grant 2009SGR 636 from the Generalitat de Catalunya to the Grup de Biologia Evolutiva is also gratefully acknowledged.

References


Received 28 April 2011; accepted 11 August 2011
Handling Editor: Charles Fox

Supporting Information

Additional Supporting information may be found in the online version of this article.

**Appendix S1.** MATLAB code ‘Gompertz.m’.

**Appendix S2.** R code ‘Gompertz.r’.

**Appendix S3.** Visualizing the effect of stochastic hazards on heat tolerance estimates.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.