Comment on 'Ecologically relevant measures of tolerance to potentially lethal temperatures'

In a recent issue of *The Journal of Experimental Biology*, Terblanche et al. (Terblanche et al., 2011) published a scholarly revision of assays to estimate critical thermal limits, and a criticism of our theoretical model explaining why estimates of upper thermal limits [CT_{max}, defined as 'the maximum temperature that an organism might potentially tolerate given its physiological condition in the absence of any other hazard' (see Santos et al., 2011)] are sensitive to rates of temperature increase (Rezende et al., 2011). Some issues have been superseded by our expansion of the original model (Santos et al., 2011), which incorporates a time-dependent survival probability function that varies with temperature to generalize what happens during a heat resistance assay, and shows that what researchers measure (knockdown temperature or time) and what they attempt to measure (CT_{max}) can differ substantially. Here we focus mainly in the problem of 'measurement' raised in their review [p. 3714 in Terblanche et al. (Terblanche et al., 2011)].

Terblanche et al. (Terblanche et al., 2011) acknowledge that 'measurement conditions may influence experimental outcomes', and suggest that some protocols are more adequate than others. However, if estimates of thermal tolerance obtained by two different methods differ, the crucial point is, why do they differ? This issue should be addressed before judging that one estimate is right, or 'ecologically relevant', and the other is not. What conditions change during the measurement? What is the attribute we want to measure? How is the best way to measure it? What is the measurement accuracy of our estimate? Are the statements on what we measure empirically meaningful regarding what we want to measure? These questions are pertinent because Terblanche et al. (Terblanche et al., 2011) also suggest that different methodologies may measure different attributes of what we call 'thermal limits', which raises the issue of which and how many traits ultimately determine thermal tolerance in a broad sense. And, more importantly, how can one assess whether these multiple (undefined) traits are being estimated correctly?

In the area of measurement accuracy it is important to distinguish between validity and reliability. 'Reliability is the agreement between two efforts to measure the same trait through maximally similar methods. Validity is represented in the agreement between two methods to measure the same trait through maximally different methods' (Campbell and Fiske, 1959). Validity is concerned with measurement bias, and reliability with measurement variance. Bias can be estimated as:

$$bias = E(X) - \Phi , \qquad (1)$$

where $E(X)$ is the expected value of the measurement over repetitions of the measurement procedure and Φ is the true value. If CT_{max} is the trait we want to measure, then it is possible to estimate the bias introduced by different protocols, and we showed that heat tolerance estimates are necessarily downwardly biased because CT_{max} is by definition the physiological limit (Santos et al., 2011). If different methods measure different traits, as suggested by Terblanche et al. (Terblanche et al., 2011), then we run into problems because we do not know what we are measuring and the concept of validity crashes.

This is more than a philosophical problem. Thermotolerance assays currently involve a myriad of conditions; for example, heating rates of 0.05, 0.06, 0.1, 0.12, 0.25 and 0.5°C min^{-1} and startup temperatures of 20, 25, 28, 35, 38 and 41°C (Terblanche et al., 2007; Chown et al., 2009; Mitchell and Hoffmann, 2010; Sgrò et al., 2010), ramping protocols that stabilize at a given temperature (Mitchell and Hoffmann 2010; Sgrò et al., 2010), or treatments involving gradual cooling followed by gradual heating (Overgaard et al., 2011a; Overgaard et al.,

2011b). If each protocol measures something different than the other, it becomes impossible to validate these measurements. Conversely, if they measure the same attribute, it remains unclear why estimates generally differ. This illustrates the fundamental pitfall of the rationale behind 'ecological realism' in the present context: it blurs the distinction between what is being measured and how it is measured.

This problem has impeded authors and referees to cross-validate estimates and dismiss results that must involve measurement error of some sort, as inconsistencies in the literature highlight. Employing a ramping assay starting at 41°C and heating rates of 0.25°C min^{-1}, Terblanche et al. [(Terblanche et al. 2007) their fig. 1] reported a knockdown temperature of 45°C after more than 25 min, which is impossible because it would take 16 min for temperatures to reach 45°C (other estimates in this figure seem to be incorrect as well). Average knockdown times of 147.9 and 158.4 min reported by Mitchell and Hoffmann [(Mitchell and Hoffmann 2010) their table 2] for *Drosophila melanogaster* fall approximately 11 standard deviation units off the 188.2 ± 3.1 min reported by Sgrò et al. (Sgrò et al., 2010) with comparable populations and the same ramping protocol (population means employed for calculations were interpolated from their fig. 1e). Importantly, Sgrò et al. (Sgrò et al., 2010) employed a starting temperature of 28°C, and not 25°C as stated in the original paper (C. Sgrò, personal communication). Similarly, hardening during fast ramping was entirely ignored in Sgrò et al. (Sgrò et al., 2010) because tolerance was sometimes expressed in units of time and in others in units of temperature (see Santos et al., 2011).

These examples illustrate that measurement validity and reliability should be a major concern, especially in the light of the growing number of counterintuitive results that have been described in recent literature. Methodology has an impact not only on mean estimates of thermal tolerance [often in opposite directions (e.g. Chown et al., 2009; Chidawanyika and Terblanche, 2011)], but also on the sign of latitudinal clines (Sgrò et al., 2010) and on estimations of phenotypic and genetic variances (Chown et al., 2009; Mitchell and Hoffmann, 2010). Our analyses suggest that these patterns can be explained on theoretical grounds and, more importantly, that some of these empirical results and others in the literature – such as the absence of correlation between estimates obtained with different methods – possibly reflect methodological artifacts (Rezende et al., 2011; Santos et al., 2011). We agree that some methodological approaches are more adequate than others, but primarily because they provide valid and reliable measures of thermal tolerance. Thus, only after understanding how measurement affects parameter estimation should one judge the relevance of these measurements from an ecological and evolutionary perspective.

References

Campbell, D. T. and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* **56**, 81-105.

Chidawanyika, F. and Terblanche, J. S. (2011). Rapid thermal responses and thermal tolerance in adult codling moth *Cydia pomonella* (Lepidoptera: Tortricidae). *J. Insect Physiol.* **57**, 108-117.

Chown, S. L., Jumbam, K. R., Sørensen, J. G. and Terblanche, J. S. (2009). Phenotypic variance, plasticity and heritability estimates of critical thermal limits depend on methodological context. *Funct. Ecol.* **23**, 133-140.

Mitchell, K. A. and Hoffmann, A. A. (2010). Thermal ramping rate influences evolutionary potential and species differences for upper thermal limits in *Drosophila*. *Funct. Ecol.* **24,** 694-700.

Overgaard, J., Hoffmann, A. A. and Kristensen, T. N. (2011a). Assesing population and environmental effects on thermal resistance in *Drosophila melanogaster* using ecologically relevant assays. *J. Thermal Biol.* **36,** 409-416.

Overgaard, J., Kristensen, T. N., Mitchell, K. A. and Hoffmann, A. A. (2011b). Thermal tolerance and tropical *Drosophila* species: does phenotypic plasticity increase with latitude? *Am. Nat.* **178,** S80-S96.

Rezende, E. L., Tejedo, M. and Santos, M. (2011). Estimating the adaptive potential of critical thermal limits: methodological problems and evolutionary implications. *Funct. Ecol.* **25**, 111-121.

Santos, M., Castañeda, L. E. and Rezende, E. L. (2011). Making sense of heat tolerance estimates in ectotherms: lessons from *Drosophila*. *Funct. Ecol.* **25**, 1169-1180.

Sgrò, C. M., Overgaard, J., Kristensen, T. N., Mitchell, K. A., Cockerell, F. E. and Hoffmann, A. A. (2010). A comprehensive assessment of geographic variation in heat tolerance and hardening capacity in populations of *Drosophila melanogaster* from eastern Australia. *J. Evol. Biol.* **23**, 2484-2493.

Terblanche, J. S., Deere, J. A., Clusella-Trullas, S., Janion, C. and Chown, S. L. (2007). Critical thermal limits depend on methodological context. *Proc. R. Soc. Lond. B* **274**, 2935-2942.

Terblanche, J. S., Hoffmann, A. A., Mitchell, K., Rako, L., Le Roux, P. C. and Chown, S. L. (2011). Ecologically relevant measures of tolerance to potentially lethal temperatures. *J. Exp. Biol.* **214**, 3713-3725.

Enrico L. Rezende* and Mauro Santos[†]

Departament de Genètica i de Microbiologia, Grup de Biologia Evolutiva (GBE),
Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain

*enrico.rezende@uab.cat
[†]mauro.santos@uab.es