Title: Temperatures that sterilise males better predict global distributions of species than lethal temperatures

Authors: Steven R Parratt[1*], Benjamin S Walsh[1], Soeren Metelmann[2], Nicola White[1], Andri Manser[1†], Amanda J Bretman[3], Ary A Hoffmann[4], Rhonda R Snook[5], Tom AR Price[1*]

Affiliations:

[1] Department of Evolution, Ecology and Behavior, University of Liverpool, Liverpool, UK.

[2] Health Protection Research Unit in Emerging and Zoonotic Infections, University of Liverpool, Liverpool, UK.

[3] Faculty of Biological Sciences, University of Leeds, Leeds, UK.

[4] School of BioSciences, Bio21 Institute, University of Melbourne, Australia.

[5] Department of Zoology, Stockholm University, Stockholm, Sweden

*Correspondence to:s.parratt@liverpool.ac.uk

†Current Address: Department of Evolutionary Biology and Environmental Studies, University of Zürich, Zürich, Switzerland

Abstract:

Climate change is well understood to be a major threat to biodiversity, but sublethal impacts of high temperatures, such as reduced fertility, have been poorly studied. We examined a panel of 43 *Drosophila* species, finding that 19 experience significant fertility loss at temperatures up to 4.3°C cooler than their lethal temperature limits. We found that upper thermal fertility limits explain global distributions of species better than limits based on lethal temperatures. This suggests that limits to reproduction, rather than limits to survival, can underpin species distributions in nature. Given that high temperatures impair male fertility across a broad range of animals and plants, many species may be at increased risk of extinction due to inability to reproduce at high temperatures.

One Sentence Summary:

Species' distributions and response to climate change are strongly affected by the temperature at which they lose fertility.

Main Text:

We urgently need to understand how rises in temperature will impact biodiversity(*1*). To do this we must understand the physiological, behavioral and evolutionary factors that underpin species' current thermal distributions(*2*, *3*). Laboratory-derived estimates of the highest temperatures at which an organism can function provide measures of species' thermal tolerances(*4*). These measures of upper thermal limits have improved the accuracy of functional species distribution models(*5*) which can be extrapolated to climate change scenarios, allowing ecologists to forecast future species distributions(*6*). Accurate predictions of species' distributions are invaluable for prioritizing conservation efforts(*7*) and predicting the invasion of disease vectors(*8*).

Upper thermal tolerance limits are usually based on the temperatures that cause loss of coordinated movement, coma, respiratory failure, or death: the species' critical thermal limit. Despite these being measured in artificial laboratory conditions, critical limits correlate reasonably well with species' distributions(*4*) and have been used to estimate species' capacity to tolerate temperature increases across their current distribution range; their 'thermal safety margins'(*3*). However, persistence of populations is not determined solely by survival, but also by reproduction. There is evidence that sub-lethal temperatures cause losses in fertility in plants(*9*), insects(*10-12*), fish(*13*), aquatic invertebrates(*14*), birds(*15*) and mammals, including humans(*16*). These effects include direct impacts on physiological processes (*10*, *15*, *17*) and indirect influences via changes in behavior and phenology(*18*). Previously, we proposed that temperatures at which fertility is lost, the thermal fertility limits (TFLs) (*18*), may be a critical

54 but understudied part of species' true upper thermal limits. If TFLs are lower than critical limits,
55 then many organisms will be more vulnerable to climate change than currently thought. If TFLs
56 and critical limits correlate poorly then we may misidentify which species are most at risk from
57 rising temperatures. Critically, we need to know whether TFLs measured in laboratory
58 conditions can predict the distribution of natural populations better than measures of critical
59 limits.
60
61 Here, we recorded three measures of upper thermal limits in adult males from 43 species of
62 *Drosophila* fruit flies. First, we exposed flies to a 4-hour static heat stress at a range of
63 temperatures from benign through to lethal(*19*). From these data we estimated the
64 temperature at which 80% of living males are sterilized (TFL80) and the temperature that is
65 lethal to 80% of individuals (LT80). Fertility was assayed at two time points: (i) over 1-6 days
66 post-heat, to capture any immediate sterilizing effect of heat, and (ii) 7-days after heat-stress to
67 capture any recovery of fertility or delayed sterility. These data allow us to compare fertility and
68 survival thresholds under identical conditions. In a separate experiment, we measured the
69 temperature at which males lose coordinated motor function under ramping temperature
70 conditions (CTmax). This measure is widely used across animal systems to predict species'
71 sensitivity to climate change (*3*, *4*).
72
73 We found that 11 of 43 species experience an 80% loss in fertility at cooler-than-lethal
74 temperatures immediately following heat-stress (Fig S1). Interestingly, rather than seeing a
75 recovery of fertility over time, the impact of high temperatures on fertility was more

76 pronounced 7-days post heat stress (Fig 1). Using this delayed measure of fertility, nearly half of
77 species (19/43) showed significant fertility loss. The difference between lethal and fertility
78 limits ranged from 0^{o}C to 4.3^{o}C (mean = 1.15 ± 0.22^{o}C), and LT80 and TFL80 predict dramatically
79 different relative ranking of species' robustness to high temperature (Fig S2).

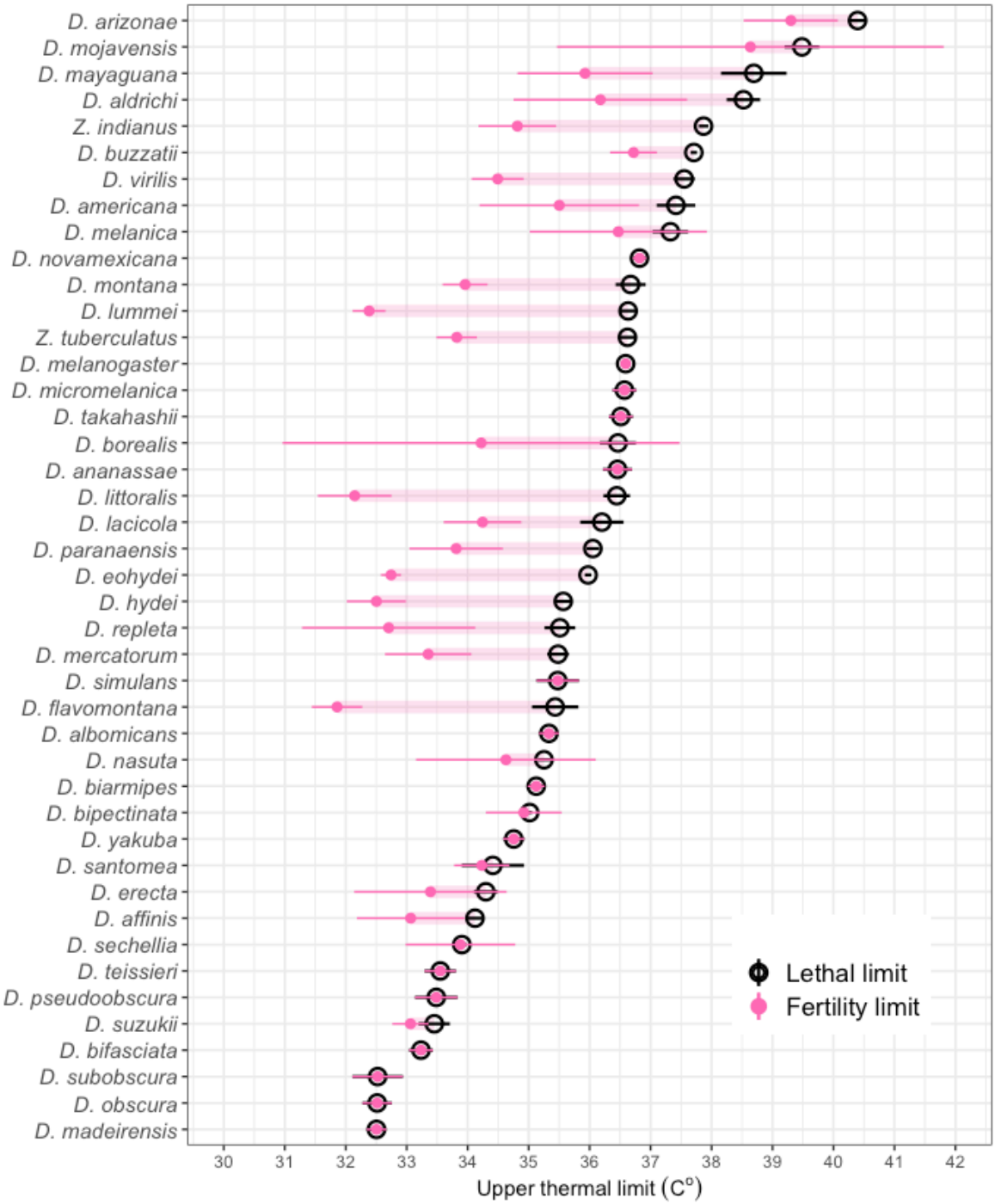


Figure 1: Incorporating fertility loss into estimates of upper thermal limits dramatically changes the ranking of species' thermal tolerance. LT80 (black circles) and TFL80 (pink

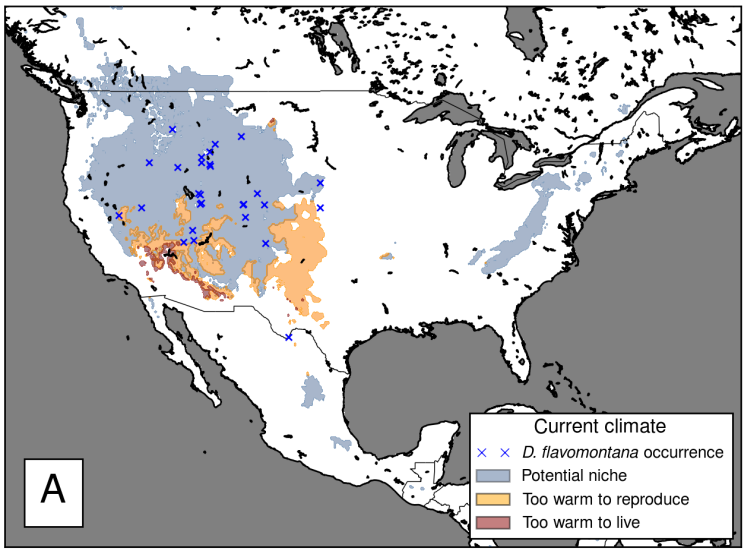
circles). 95% CI are shown as error bars for both measures. Fertility loss measured seven days post heat stress to account for delayed sterility effects.

These data show that fertility loss at sub-lethal temperatures is common across *Drosophila.* This suggests that ignoring TFLs has led to the overestimation of thermal tolerance for many species. This may be particularly problematic for efforts to preserve rare and endangered species or control emerging pests and disease vectors. For example, we found that the crop pest species *Zaprionus indianus* (African fig fly) has a relatively high upper lethal limit (LT80 = 37.9^{o}C) but a much cooler upper fertility limit (TFL80 = 34.8^{o}C). This switches the species' thermal hardiness rank from rank 5th to 17th of species here (Fig S2). This pest is currently expanding throughout the New World, and this disparity could have marked consequences for forecasting its putative range and hence where to concentrate control efforts(*20*).

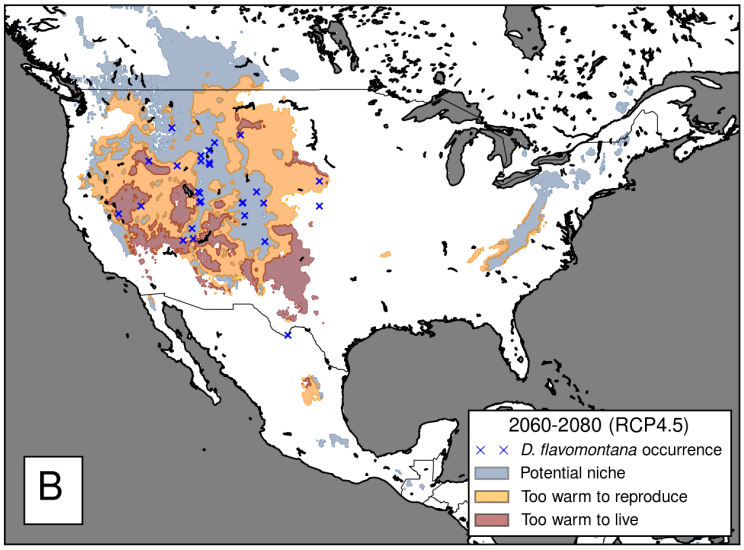
Our work demonstrates that TFLs can be substantially lower than lethal temperatures when measured under the same conditions in the laboratory. However, the key question is whether TFLs actually shape organisms' distributions in nature. To test this, we used existing data on the distribution of each *Drosophila* species and integrated this with global climate data for each location to estimate the maximum temperatures species are likely to encounter in nature(*19*). Previous work using 95 *Drosophila* species found that the mean maximum summer temperature measured in this way correlates with CTmax for species from dry (<1000ml annual rainfall) but not from wet (>1000ml annual rainfall) environments(*4*). We first verified this finding in our smaller dataset of 43 species; CTmax significantly interacts with annual

105 precipitation to predict mean maximum environmental temperature (PGLS: $F = 6.02_{1,39}$, $P =$
106 0.018, $adjR^2 = 0.37$). We then tested whether either LT80 or TFL80 measured 7-days after heat
107 stress are better predictors of species' thermal habitats than CTmax (Table S1). Across all
108 species, we found that a model with TFL80 as the sole physiological predictor had the strongest
109 correlation with mean maximum environmental temperature with no significant interaction
110 with precipitation (PGLS: interaction term: $F = 0.531_{1,39}$ $P = 0.471$, TFL80: $F = 39.3_{1,39}$ $P < 0.001$,
111 $adjR^2 = 0.55$). LT80 also did not significantly interact with precipitation to predict environmental
112 temperature but explained considerably less variation than TFL80 (PGLS: interaction term: $F_{1,39}$
113 $= 0.99$, $P = 0.324$, LT80: $F_{1,39} = 0.99$, $P = 0.324$, $adjR^2 = 0.31$). Hence, TFL results in a 48% and
114 77% improvement in the accuracy (R^2) of species distribution predictions compared to CTmax
115 and LT80 respectively. This dramatic improvement is largely due to TFL80 more accurately
116 predicting the thermal range for species from high humidity environments (Table S2 & Fig S3).
117 We find qualitatively similar results if both lethal and sterilizing limits are included as predictors
118 in the same model (Table S3). The power of TFL80 to predict the distribution of *Drosophila* in
119 nature strongly suggests that fertility losses due to high temperature are an important, but
120 previously ignored, determinant of where species can persist.
121
122 Predicting species' vulnerability to current and future temperature extremes is critical to
123 protect biodiversity(*1*). "Thermal safety margins" measure species' vulnerability to climate
124 change by taking the difference between an organism's physiological thermal limit and the
125 maximum air temperature it is likely to experience(*21*). Although safety margins often use
126 complex measures of environmental temperature to improve accuracy (e.g. microhabitat

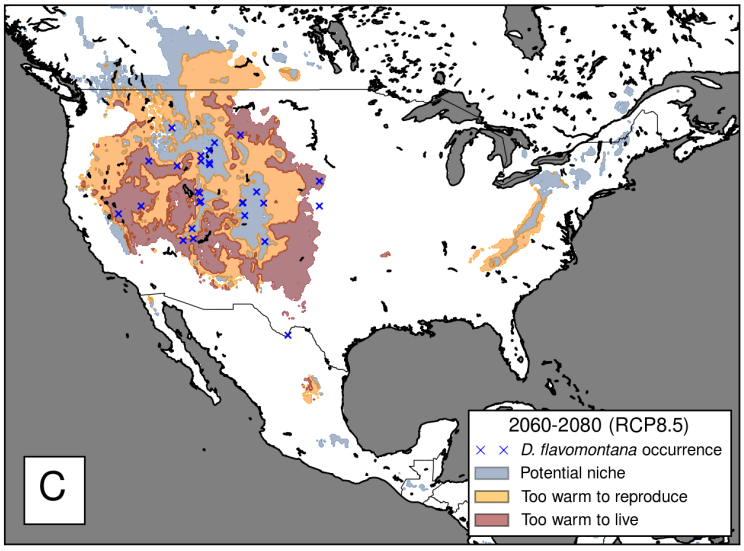
127 availability, thermoregulation behavior of adults(*3*)), they remain reliant on relatively simple
128 measures of thermal vulnerability. We find that TFL80 measured seven days after heat-stress
129 produces significantly smaller safety margins than either LT80 or CTmax (LMER: χ^2 = 40.36, df =
130 2, $P < 0.001$, Fig S4). For populations that already live near the upper edge of their thermal
131 range, TFL-based safety margins are reduced from 2.59 ± 3.08°C under assumptions of LT80, to
132 1.43 ± 2.80°C. This predicts that 20 of the 43 species studied here are found in environments in
133 which air temperatures exceed safety limits during the hottest part of the year. We illustrate
134 the implications of this with case-study distribution models of *Drosophila flavomontana;* safety
135 margins based on TFL80 predicts a 17.9% reduction in habitable landscape compared to an
136 identical LT80-based model under current climate conditions (Fig 2A). This disparity between
137 predictions based on sterility and lethality grew to 48.0% by the year 2070 under moderately
138 optimistic future climate forecasts (RCP4.5, Fig 2B), and to 58.9% under pessimistic climate
139 change scenarios (RCP8.5, Fig 2C). TFL-based models also predict that by 2070 the available
140 habitat for *D. flavomontana* will have reduced by 42.3% and 62.9% under RCP4.5 and RCP8.5
141 respectively.



142



143



144

Figure 2: Potential habitat range of *Drosophila flavomontana* (LT80 = 35.8°C, TFL80 = 31.4°C) under A) current and B & C) possible future climate scenarios (B = RCP4.5 'moderately optimistic', C = RCP8.5 'pessimistic', predicted for 2060 - 2080). Colored areas in each panel represent suitable habitat range predicted by the model that excludes maximum temperature (Fig S5). Red areas show regions where maximum summer temperatures exceed LT80. Orange areas show regions where maximum summertime temperatures exceed TFL80. Blue regions are habitable for *D. flavomontana* all year.

If our data for *Drosophila* can be extrapolated to other organisms, it is likely that male fertility losses at high temperatures are common and can occur at substantially lower temperatures than the species upper lethal limit. The limited data on fertility at extreme temperatures supports this, with losses in male fertility at high temperatures seen across a broad diversity of organisms(*18*). It is possible that behavioral thermoregulation will reduce the impact of high temperatures on fertility in nature. *Drosophila* are able to behaviorally thermoregulate, moving to leaf litter, shade, or higher altitudes, with many species able to survive high temperature periods by aestivating as adults, eggs or pupae(*22*). Despite this, we still find that the distribution of species is predicted by thermal fertility limits. Even species that might be predicted to have high thermal tolerance can show evidence of thermal fertility losses. For instance, the zebra finch, a high temperature adapted desert- dwelling endothermic species with naturally high body temperature and good thermoregulation, shows substantial damage to sperm at temperatures it regularly experiences in nature(*15*).

Our work suggests that temperature-driven fertility losses may be a major threat to biodiversity during climate change. We urgently need to understand the range of organisms likely to suffer similar fertility losses in nature, and the traits that predict vulnerability. However, we currently do not understand the physiology underlying variation in TFLs between species, nor the selective forces that created this variation. Ultimately, we need to know whether evolution for higher TFLs will allow species to adapt to a warming environment.

References and Notes:

1. G. T. Pecl *et al.*, Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science*. **355**, eaai9214 (2017).

2. I.-C. Chen, J. K. Hill, R. Ohlemüller, D. B. Roy, C. D. Thomas, Rapid Range Shifts of Species Associated with High Levels of Climate Warming. *Science*. **333**, 1024–1026 (2011).

3. J. M. Sunday *et al.*, Thermal-safety margins and the necessity of thermoregulatory behavior across latitude and elevation. *Proceedings of the National Academy of Sciences*. **111**, 5610–5615 (2014).

4. V. Kellermann *et al.*, Upper thermal limits of Drosophila are linked to species distributions and strongly constrained phylogenetically. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16228–16233 (2012).

5. M. Kearney, W. Porter, Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecol Lett*. **12**, 334–350 (2009).

6. J. Elith, J. R. Leathwick, Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annu. Rev. Ecol. Evol. Syst.* **40**, 677–697 (2009).

7. A. J. Nowakowski *et al.*, Thermal biology mediates responses of amphibians and reptiles to habitat modification. *Ecol Lett*. **21**, 345–355 (2018).

8. S. Metelmann *et al.*, The UK's suitability for Aedes albopictus in current and future climates. *Journal of The Royal Society Interface*. **16**, 20180761 (2019).

9. T. L. Sage *et al.*, The effect of high temperature stress on male and female reproduction in plants. *Field Crops Research*. **182**, 30–42 (2015).

10. K. Sales *et al.*, Experimental heatwaves compromise sperm function and cause transgenerational damage in a model insect. *Nat Commun*. **9**, 4771 (2018).

11. D. Porcelli, K. J. Gaston, R. K. Butlin, R. R. Snook, Local adaptation of reproductive performance during thermal stress. *J Evol Biol*. **30**, 422–429 (2016).

12. A. D. Saxon, E. K. O'Brien, J. R. Bridle, Temperature fluctuations during development reduce male fitness and may limit adaptive potential in tropical rainforest Drosophila. *J Evol Biol*. **31**, 405–415 (2018).

13. R. D. Breckels, B. D. Neff, The effects of elevated temperature on the sexual traits, immunology and survivorship of a tropical ectotherm. *Journal of Experimental Biology*. **216**, 2658–2664 (2013).

14. C. W. Paxton, M. V. B. Baria, V. M. Weis, S. Harii, Effect of elevated temperature on fecundity and reproductive timing in the coral Acropora digitifera. *Zygote*. **24**, 511–516 (2016).

15. L. L. Hurley, C. S. McDiarmid, C. R. Friesen, S. C. Griffith, M. Rowe, Experimental heatwaves negatively impact sperm quality in the zebra finch. *Proc. R. Soc. B.* **285**, 20172547 (2018).

16. L. Yogev *et al.*, Seasonal variations in pre- and post-thaw donor sperm quality. *Hum Repro*. **19**, 880–885 (2004).

17. C. Rohmer, J. R. David, B. Moreteau, D. Joly, Heat induced male sterility in Drosophila melanogaster: Adaptive genetic variations among geographic populations and role of the Y chromosome. *Journal of Experimental Biology*. **207**, 2735–2743 (2004).

18. B. S. Walsh *et al.*, The Impact of Climate Change on Fertility. *Trends Ecol. Evol*. **34**, 249–259 (2019).

19. Materials and Methods

20. K. van der Linde, G. J. Steck, K. Hibbard, J. B. Florida, First records of Zaprionus indianus (Diptera: Drosophilidae), a pest species on commercial fruits from Panama and the United States of America. *BioOne* **89,** 402 - 404 (2006).

21. C. A. Deutsch *et al.*, Impacts of climate warming on terrestrial ectotherms across latitude. *Proceedings of the National Academy of Sciences*. **105**, 6668–6672 (2008).

22. M. E. Dillon, G. Wang, P. A. Garrity, R. B. Huey, Thermal preference in Drosophila. *Journal of Thermal Biology*. **34**, 109–119 (2009).

23. J. Overgaard, M. R. Kearney, A. A. Hoffmann, Sensitivity to thermal extremes in Australian Drosophila implies similar impacts of climate change on the distribution of widespread and tropical species. *Global Change Biology*. **20**, 1738–1750 (2014).

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

25. P. T. Rohner *et al.*, Interrelations of global macroecological patterns in wing and thorax size, sexual size dimorphism, and range size of the Drosophilidae. *Ecography*. **41**, 1707–1717 (2018).

26. S. E. Fick, R. J. Hijmans, WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*. **37**, 4302–4315 (2017).

27. S. J. Phillips, R. P. Anderson, R. E. Schapire, Maximum entropy modeling of species geographic distributions. *Ecological Modelling*. **190**, 231–259 (2006).

28. N. Ramankutty, A. T. Evan, C. Monfreda, J. A. Foley, Geographic distribution of global agricultural lands in the year 2000, *Global Biogeochemical Cycles*, **22** doi:10.1029/2007GB002952, (2008).

29. C. Amante, B. W. Eakins, ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Source and Analysis. *National Geographic Data Center* (2009).

30. C. Liu, M. White, G. Newell, Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography*. **40**, 778–789 (2013).

31. B. Thrasher, E. P. Maurer, C. McKellar, P. B. Duffy, Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrol. Earth Syst. Sci.* **16**, 3309–3314 (2012).

32. A. E. Jones *et al.*, Bluetongue risk under future climates. *Nature Climate Change*. **9**, 153–157 (2019).

33. S. Pitnick, T. A. Markow, G. S. Spicer, Delayed male maturity is a cost of producing large sperm in Drosophila. *Proceedings of the National Academy of Sciences*. **92**, 10614–10618 (1995).

34. K. E. Roberts, J. D. Hadfield, M. D. Sharma, B. Longdon, Changes in temperature alter the potential outcomes of virus host shifts. *PLoS pathogens*. **14**, e1007185 (2018).

35. M. D. Garlovsky, R. R. Snook, Persistent postmating, prezygotic reproductive isolation between populations. *Ecol Evol*. **8**, 9062–9073 (2018).

36. A. A. Hoffmann, S. L. Chown, S. Clusella Trullas, Upper thermal limits in terrestrial ectotherms: how constrained are they? *Funct Ecol*. **27**, 934–949 (2013).

Acknowledgements: Natasha Mannion and Dr Angela Sims for assistance with experiments, Dr Ben Longdon and Dr Katherine Roberts for flies, Rowan Connell for designing 3D-printed equipment. **Funding:** NERC grant NE/P002692/1 to TP, AB, AH & RS. SNF P300PA_177830 to AM. NIHR HPRU EZI to SM. **Author Contributions:** Conception: TP, AB, RS, AH, SP. Methodology & data collection: SP, BW, NW. Data curation: SP & SM. Analysis: SP, AM & SM. Original Draft: SP, TP, AB, RS & AH. Review & Editing: All authors. **Data and materials availability:** All data and analysis R code will be deposited on Dryad upon acceptance of this manuscript.

List of Supplementary Materials:

Materials and Methods

Figures S1-S5

Tables S1-S4

Supplementary Text

References (23 - 36)

277 Materials and Methods:

278 Animal maintenance:

279 All species were kept in temperature-controlled rooms at their noted "rearing temperature"
280 (see table S4) selected based on observations of when the laboratory populations were most
281 stable. All stocks were kept at 12:12 L:D and ambient humidity. Stocks were maintained at
282 moderate density (20 – 50 flies per 23ml vial or 50 – 100 flies per 300ml bottle culture).
283 Cultures were tipped to fresh food every 7 days and a new generation was made every 4 – 6
284 weeks depending on the speed of a species development. Species were reared on one of four
285 food types; A = ASG (10g agar, 85g Sucrose, 20g yeast extract, 60g maize, 1000ml H_2O, 25ml
286 10% Nipagin), B = Banana (10g agar, 30g Yeast extract, 150g pulped banana, 50g Molasses, 30g
287 Malt Extract, 25ml 10% Nipagin, 1000ml H_2O), P = Propionic (10g Agar, 20g Yeast extract, 70g
288 cornmeal, 10g soya flour, 80g Malt Extract, 22g Molasses, 14ml 10% nipagin, 6ml Propionic
289 acid, 1000ml H_2O), M = Malt (10g agar, 20g Yeast, 60g Semolina, 80g Malt, 14ml 10% Nipagin,
290 5ml Propionic acid, 1000ml H_2O). To standardize across species, we only included males in our
291 TFL assay that were sexually mature. This inevitably led to variation in the age of males (days
292 since eclosion) between species based on published age-to-maturity (e.g. *34*) and our own
293 observations (Table S4 for details).

294 Species verification:

295 Most *Drosophila* species were gained from stock centers, other research groups, or through
296 field collections. Where species were not distinguishable through evident morphological traits,
297 we verified position within the phylogeny by nucleotide sequence identity.

298

299 We extracted DNA from 2-3 adult male flies with DNeasy kits (Qiagen) following the
300 manufacturer's invertebrate protocol. We PCR amplified a portion of the mitochondrial
301 universal barcode gene cytochrome oxidase subunit 1 using the primers C1-J-1718 (5' –
302 GGAGGATTTGGAAATTGATTAGT – 3') and C1-N-2191 (5' – CCCGGTAAAATTAAAATATAAACTTC –
303 3') using HotStart Taq (Promega) with (5-minute initial heating, 30 cycles at 95°C for 30s, 56 for
304 30s, and 72°C for 30, with an final elongation step of 72°C for 120s). PCR products were
305 visualized by SYBRSafe-stained gel electrophoresis and cleaned up using Exonuclease I and
306 Shrimp Alkaline Phosphatase incubation per supplier protocol (BioLine). We then used BigDye
307 based sequence reactions with both forward and reverse primers, followed by NaOH & ethanol
308 precipitation clean-up and precipitation before sequences were analyzed with an ABi 3500XL
309 Genetic Analyzer. Forward and reverse sequences for each species were aligned to derive a
310 consensus sequence. For species with a putative ID based on stock center or collaborator
311 expertise we assessed the clustering of these sequences with publicly available CO1 sequences
312 from the same species available on the BOLD database (boldsystems.org). Clustering was
313 performed with the MAFFT server tool, and clustering via a neighbor-joining tree was
314 visualized, all based on default parameters. For species that were unknown (two *repleta*-like
315 species collected from the field in Madeira and Morocco), we performed the same process as

316 above, but included multiple sequences from multiple species in the *repleta* species group –
317 initial BLAST of these sequences on the NCBI database had given an approximate idea of the
318 species ID.
319
320 Measuring upper thermal limits
321 We assayed three metrics of upper thermal limits in males from 43 species of *Drosophila*: Lethal
322 Temperature 80% (LT80), Thermal Fertility Limit 80% (TFL80) and Maximum Critical
323 Temperature (CTmax). We chose to test for thermal fertility limits in only males because of the
324 substantial evidence that male fertility is particularly vulnerable to high temperatures(*18*).
325
326 *LT80 & TFL80:* Newly eclosed virgin adult males were collected over a 48-hour period into
327 *Drosophila* vials and allowed to sexually mature for 7 – 21 days depending on species (see Table
328 S4). Males were kept in groups of maximum 10 individuals to minimize intrasexual aggression.
329 Once mature, males were transferred to fresh 23ml fly vials with standard maize-sugar-yeast
330 agar "ASG" medium' food and allocated to 3D-printed floating plastic racks in pre-heated
331 waterbaths (Grant TXF200) set to a range of temperatures (see Table S4 for temperature
332 ranges). ASG food was used during heating for all species because preliminary work had shown
333 that survival under heat stress was influenced by food type. Floating racks were weighted with
334 ball-bearings at each corner to keep vials containing flies submerged so that the waterline was
335 above the cotton-wool stopper. This ensured that flies were exposed to homogenous
336 temperature conditions inside the vials. In each run of the experiment we monitored the
337 temperature inside a fresh ASG vial without flies with a k-type thermocouple attached to a data

logger (Pico Technology TC-08). Males were heated for 4 hours between ~10am - ~2pm and then returned to temperature-controlled rooms set to the species' benign temperature. We scored the survival of males the next morning to account for any immediate recovery or delayed lethality of the temperature treatment. We scored flies as dead if they were completely immobile or had become stuck on their backs in the fly food – the latter would indicate that the flies were completely incapacitated by the treatment temperature which would likely be lethal in the wild.

Surviving males were then individually aspirated into separate, fresh vials containing their species' designated food type and 3-4 sexually mature virgin females. Males remained in these vials at their benign temperature and allowed to mate freely for 6 days. This allows us to score any immediate lasting effect of temperature on fertility. After 6 days males were transferred to a second fresh vial with 1-2 more virgin females and allowed to mate for 24 hours. This allowed us to test for recovery of fertility or any delayed sterilizing effect of temperature. Females in both sets of vials were left to lay eggs for 3-5 days (depending on fecundity of the species) after the male had been removed. Fertility was scored qualitatively as the presence or absence of any offspring: vials were routinely scored for the presence of larval flies by checking for the larvae themselves or their distinctive track marks in the food. Juveniles flies were allowed to develop through to adult eclosion before being frozen and the number of adult offspring counted.

In total, 14742 males were exposed to heat treatment, of which 10925 survived and 9064 went on to be tested for heat-induced sterility. The average sample size at each temperature per species was 36.5 ± 0.5sem, ranging from 10 – 88 individuals.

CTmax: The measures of upper tolerance described above are performed under static heat-stress conditions – this is necessary for estimating TFLs because the phenotype "fertility" can only be observed after the heat-stress treatment, so matching a male's fertility to a stress temperature requires that he only experiences a single stressful environment. LT80 is therefore a directly comparable measure to TFL80. However, several studies in *Drosophila* and other animals have used estimates of temperature tolerance measured under gradually ramping heat conditions to estimate species response to climate change (*4*, *7*, *23*, *24*). Because of this, and because previous work has shown that estimate of upper critical thermal limits can depend on methodological context(*25*), we also measured upper critical limits of our 43 *Drosophila* species under ramping heat conditions – critical thermal maxima (CTmax). This allows us to compare how static and dynamic measures of critical heat tolerance predict species' distributions. It also allows us to compare upper TFLs to two widely used measures of upper critical limits.

We measure CTmax as the mean temperature at which male flies lose coordinated movement and are unable to right themselves. Individual males were collected as virgin adults from fly stocks and allowed to sexually mature (see Table S4). Flies were then anesthetized with CO_2 and isolated into individual glass sample tubes sealed with a rubber stopper. Sample tubes were left at ambient room temperature for at least 30 minutes to allow males to fully recover from CO_2 anesthetization – preliminary trials found no difference in CTmax between anesthetized and non-anesthetized male *D. melanogaster.* Sample tubes then were attached to a custom-built plastic rack and held at a 45^o angle with elastic banding to facilitate the

382 experimenter identifying when flies collapsed. The rack and vials were submerged in a glass
383 aquarium held at the rearing temperature for the given species by a water circulator (Grant
384 TXF200) (18, 23 or 25^{o}C – Table S4). The temperature was then increased at a rate of 0.1^{o}C per
385 minute until all flies were incapacitated. The temperature at which flies collapsed for 30
386 seconds and did not right themselves after being physically disturbed by the experimenter
387 (tapping the vial) was scored. The temperature of the waterbath was continuously recorded
388 throughout the assay with a data logger (Elitech RC-61). The mean sample size for each species
389 for CTmax assays was 19.53 ± 1.77.
390
391 Repeatability
392 Due to the logistical constraints in the number of flies that could be processed at a time and the
393 differences in temperature range required to stress each species it was impossible to assay
394 upper limits across all 43 species simultaneously. Further, inter-species variation in
395 development time prohibited us from synchronizing all 43 species to run the assay in a
396 completely randomized block design. Instead, we assayed each species separately to
397 independently estimate CTmax, TFL80 and LT80. These assays were run between June 2018 –
398 November 2019.
399
400 To test the repeatability of our measure of Lethal Temperature (LT80) we ran a single control
401 block in February 2020 with 22 species that we could obtain enough virgin males that would
402 reach sexual maturity in synchrony. We simultaneously exposed these species to a range of 6
403 temperatures each that covered their estimated LT80 identified when we had previously run

404 them individually. We then tested the strength of the correlation between our original LT80
405 estimates and those given by this replicate block and determined if any given species showed
406 significantly different estimates based on whether the confidence intervals of the two runs
407 overlapped. Sample sizes for this control run ranged from 3 – 10 males per temperature per
408 species, and involved 1152 male flies in total.
409
410 As a test of repeatability for our TFL80 estimates we ran two independent TFL assays on
411 *Drosophila virilis*. This species was selected because it had demonstrated a clear difference
412 between TFL80 and LT80 limits. The two repeat blocks were run 6 months apart and data were
413 collected by two separate researchers. We compare the estimated TFL80 from these
414 independent runs and consider differences to be significant if the 95% confidence intervals of
415 our estimates do not overlap.
416
417 Repeatability of CTmax was estimated by randomly allocating species from the same rearing
418 temperatures into blocks of 3 - 5 species and measuring their mean CTmax as described above.
419 We compared these mean Ctmax values with those obtained when species were run
420 individually when we collected the CTmax data used in analyses.
421
422 <u>Point estimates of TFL80 & LT80</u>
423 We chose 80% thresholds as our physiological limits for both LT and TFL because i) we wanted
424 to compare viability and fertility measures on comparable scales, and ii) a loss of 80% of males
425 is likely to represent a substantial threat to population stability for most species.

We found that general linear models fit our qualitative (1/0 data) fertility and survival data poorly, and overestimated 80% fertility loss thresholds considerably. Instead, we generated point estimates of 80% thresholds with dose response models (`drm()`), implemented in the `drc` package in R. We used 3-parameter versions of the log-logistic model which fixes the lower limit of the predicted curve to 0 to reflect that we expect fertility and survival to eventually reach 0 over an infinite range of temperatures. The models then estimate the point at which we see an 80% reduction in survival or fertility relative to the upper limit present in the data. In this way, these models allow for a given species to inherently have some degree of sterility even at benign temperatures, and to estimate the temperature required to reduce this by 80%. Percentage fertility was analyzed for surviving flies only, i.e males killed by the heat treatment were not considered to be sterile.

In analyses in which we predict geographic range by physiological limits we use TFL80 estimates that are 'significantly' lower than the species' LT80 estimate – for the remaining species TFL80 point estimates are defaulted to the LT80 to be conservative. We considered the difference between TFL80 and LT80 to be 'significant' if the 95% confidence intervals of the two point-estimates do not overlap.

Correlations between upper thermal limits.

We used phylogenetically controlled linear models to test how much variation in TFL80 is explained by variation in LT80. These measures are taken from the same assays and are often

very similar, so this correlation gives a quantitative assessment of the mismatch between these two thermal tolerance traits.

CTmax and LT80 inherently use different heating regimes to capture upper critical limits. Most notably, CTmax includes an element of temperature ramping, and so may allow species to plastically harden to high temperature stress. To test how well these two measures capture the same component of thermal interspecific variation, and are therefore proxies for each other, we used phylogenetically controlled linear models to explore the correlation between TFL80 and CTmax. Whilst we did not expect these methods to produce identical point estimates of lethal temperatures for a given species, we did hypothesize that these correlations will have a high R^2 and a slope estimate close to 1.

If there is a universal upper limit to fertility across *Drosophila* then we would expect the difference between TFL and critical temperatures to be greater in species with higher temperature LT80 and CTmax. To test this, we used phylogenetically controlled linear models to test for significant correlations between ΔLimits (the difference between TFL80 and LT80) and both LT80 and CTmax.

<u>Accounting for phylogeny:</u>

All of the species we used in this study are present in a phylogeny published by Patrik Röhner and colleagues(*26*): We use this phylogenetic tree as the basis of all our phylogenetically controlled analyses.

We tested for phylogenetic signal in all three upper thermal limits (LT80, TFL80 and CTmax). We used `pgls()` in `caper` to estimate if Pagel's lambda was significantly different from 0 (no phylogenetic signal) or 1 (complete Brownian motion). The maximum likelihood method was used to estimate lambda.

When analyzing correlations amoung traits, and between traits and geographic distributions we corrected for phylogenetic signal in model residuals with `pgls()` in the R package `caper`. Even where phylogenetic signal was estimated to be not significantly different to 0, we retained the correction in the model because these models give qualitatively identical results to uncorrected linear models using `lm()`.

<u>Environmental variables:</u>

Geographic distribution data for each species were obtained manually from TaxoDros.uzh.ch as coordinates and location names. These were systematically cleaned by deleting any entry set to zero decimal places in both coordinates. Then, they were globally cleaned by removing locations with vague names such as entire countries or US states. Plots of each species location on a map were then visually inspected, and outliers were investigated and removed or corrected when appropriate.

The cleaned dataset of geolocations was then integrated with rasterized bioclimate data from the WordlClim V2.0 database (*27*). Our intention was not to fully reassess multiple climate

492 variables as predictors for *Drosophila* distributions, because this has received extensive
493 attention in the existing literature(*4*, *24*). Instead, we tested if TFLs correlate with variables
494 previously identified as important for predicting physiological limits and distribution in
495 *Drosophila*; the maximum temperature reached in the hottest summer month (WorldClim Bio5)
496 and annual precipitation (WorldClim Bio12). Means of each climate variable for each species
497 were calculated across all global locations. We also calculated the mean of each variable
498 independently for locations in the North and South hemispheres for species whose range
499 straddles the equator.
500
501 Comparing upper thermal limits as predictors of species' thermal environment
502 We used three distinct modelling approaches to test which measure of upper thermal limits
503 best predicts species' thermal range:
504
505 Firstly, we fitted separate phylogenetically controlled linear models with either TFL80, LT80 or
506 CTmax as independent predictors, and species' Tmax (WorldClim Bio5) as a response variable.
507 In all three models we included mean annual precipitation (WorldClim Bio12 or "Pann") as an
508 interacting covariate, because previous work has found that the power of CTmax to predict
509 species' thermal environments degrades in more humid habitats(*4*). In these models all
510 predictors are centered and scaled. We then used AIC-based model selection to reduce models
511 to minimum-adequate models, and compared the slope estimate, model structure and R^2 of the
512 competing models.
513

514 Secondly, we directly replicated the method of (*4*) in which the authors split species into those
515 that experience greater-than and less-than 1000mm of rainfall ("wet" and "dry" hereafter). The
516 authors of (*4*) found that CTmax correlated well for "dry" species but not for "wet" species. We
517 ran phylogenetically corrected linear models with Tmax as a response and either CTmax, LT80
518 or TFL80 and predictors for both 'dry' and 'wet' species. We qualitatively compared the R^2 of
519 these models as indicators of model fit.
520
521 Finally, we included both CTmax and TFL80 and their interaction with `Pann` into a single
522 phylogenetically controlled model of Tmax. We did not include LT80 in this analysis because *a*
523 *priori* Variance Inflation Factor tests (VIF) on non-phylogenetically controlled versions of this
524 model indicated that LT80 correlated very strongly with both other measures. We reduced the
525 maximal model through AIC-based stepwise simplification starting with higher order interaction
526 terms to reach a minimum adequate model.
527
528 <u>Calculating and modelling thermal safety margins</u>
529 Thermal safety margins were calculated as the difference between a species' physiological
530 upper thermal limit and the mean maximum summertime temperature they are likely to
531 experience across all of their known global distribution locations. As per(*4*), we calculated this
532 in two ways: firstly by taking the mean Tmax for each species, to estimate safety margins for
533 the 'central' populations of a species (termed "Central Safety Margin"). Secondly by taking the
534 mean Tmax + 1 SD to account for populations of species that live at the upper thermal range of

535 species' distribution (the "distribution safety margin). Safety margins were calculated
536 separately using either CTmax, TFL80 or LT80 as the physiological limit.
537
538 We also treated populations in the North and Southern hemispheres as equivalent to separate
539 species, because previous work has found minor differences in safety margins in the North and
540 South hemisphere, likely because of biases in sampling efforts(*3*).
541
542 Safety margins were modelled as phylogenetically controlled linear mixed models with
543 'hemisphere' (N or S) and 'physiological limit' (CTmax, TFL80 or LT80) as linear predictors,
544 absolute latitude (degrees^2) as a quadratic predictor, and species identity as a random
545 intercept.
546
547 Predicting current and putative future distributions of *D. flavomontana*
548 As a case-study demonstration of the importance of TFLs, we use MaxEnt v3.4.1 (*28*) to model
549 potential distributions of *D. flavomontana*. Occurrence data were obtained from
550 TaxoDros.uzh.ch and points within 25km of each other were clustered together to avoid
551 oversampling (N = 24). First, we chose climatic and other variables to predict a potential
552 distribution of *D. flavomontana* that is independent of the upper thermal limit. After excluding
553 variables that strongly correlated or had little to no impact on the model results (<3%
554 contribution), we used four of the tested 15 variables: mean winter temperature (WC Bio 11,
555 from the WorldClim2.0 database, precipitation seasonality (WC Bio 15), an agricultural index
556 (*29*), and elevation (*30*). We then ran 10 cross-validations with maximum iterations of 5000,

557 using 10000 random pseudo-absences. To distinguish suitable from unsuitable regions, we used
558 the threshold that maximized Youden's index (maximum training sensitivity plus specificity)
559 (*31*). The MaxEnt mean output grid is shown in Figure S5.
560
561 We then constrained the suitable area by the upper physiological thermal limits measured
562 above: i) the lethal temperature measured as LT80 (35.8°C), and ii) the TFL80 (31.4°C). We use
563 the maximum weekly temperature of the warmest month as reference (WC Bio5). Using WC Bio
564 05 directly as a predictor variable for *D. flavomontana* had close to no effect on the distribution
565 (0.1% contribution).
566
567 For future predictions, we chose two carbon emission scenarios: RCP4.5 (moderately optimistic)
568 and RCP8.5 (pessimistic). We use precipitation and temperature predictions of the NorESM1-M
569 model from the NEX-NASA GDDP data set (*32*), as this model produces a median trend in terms
570 of temperature increase(*33*). The agricultural index and elevation had to be used as present day
571 data. Predicting and constraining the possible future distribution followed the same procedure
572 as for the current distribution.
573

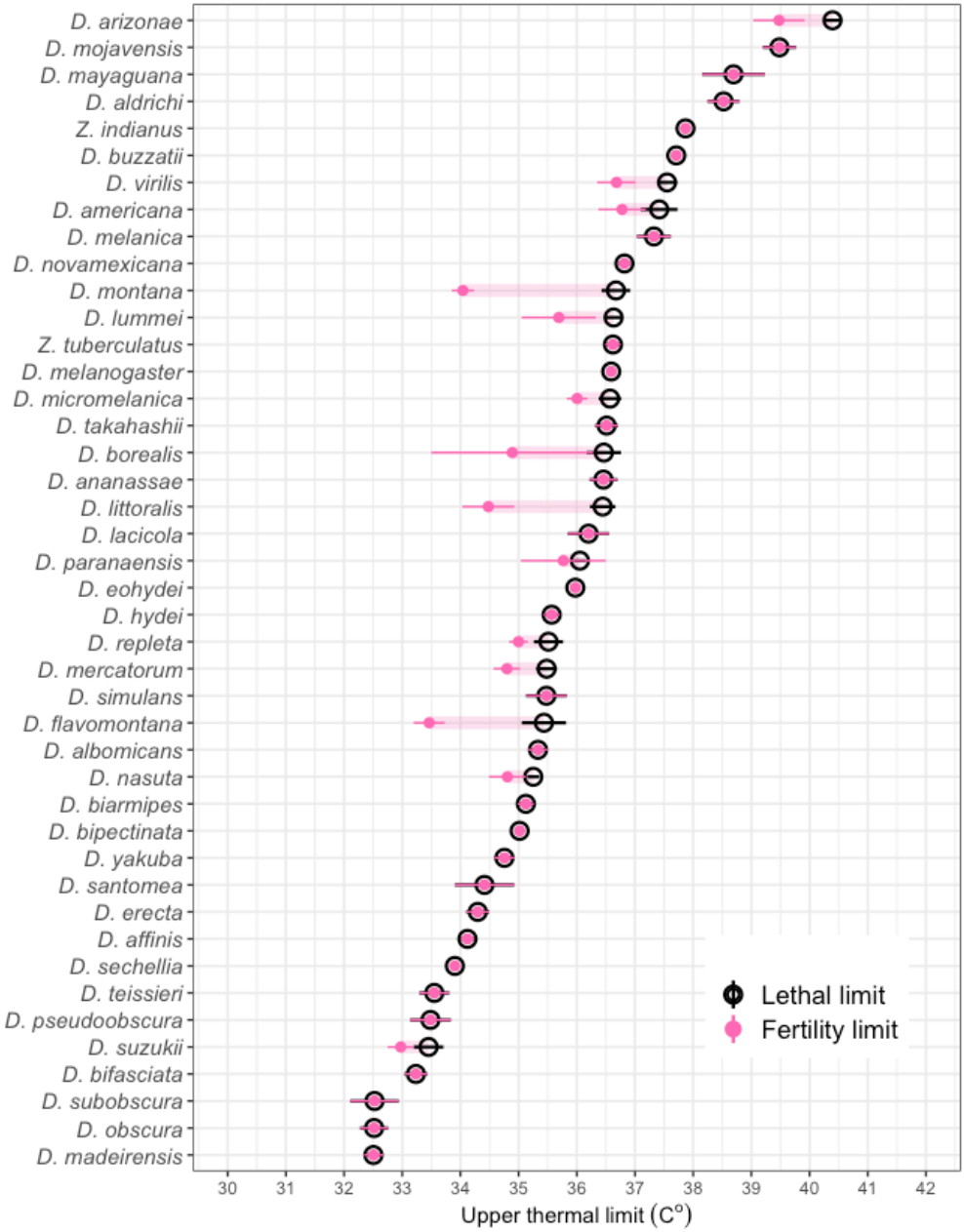


Figure S1: TFL80 and LT80 of *Drosophila* immediately after heat stress. Several species of *Drosophila* lose 80% fertility at cooler-than-lethal temperatures immediately following heat-shock. LT80 (black circles) and TFL80 (pink circles). Errors for both measures are 95% confidence intervals generated from dose response model estimates. Fertility loss measured as ability to sire any offspring between 1 – 6 days post heat-stress.

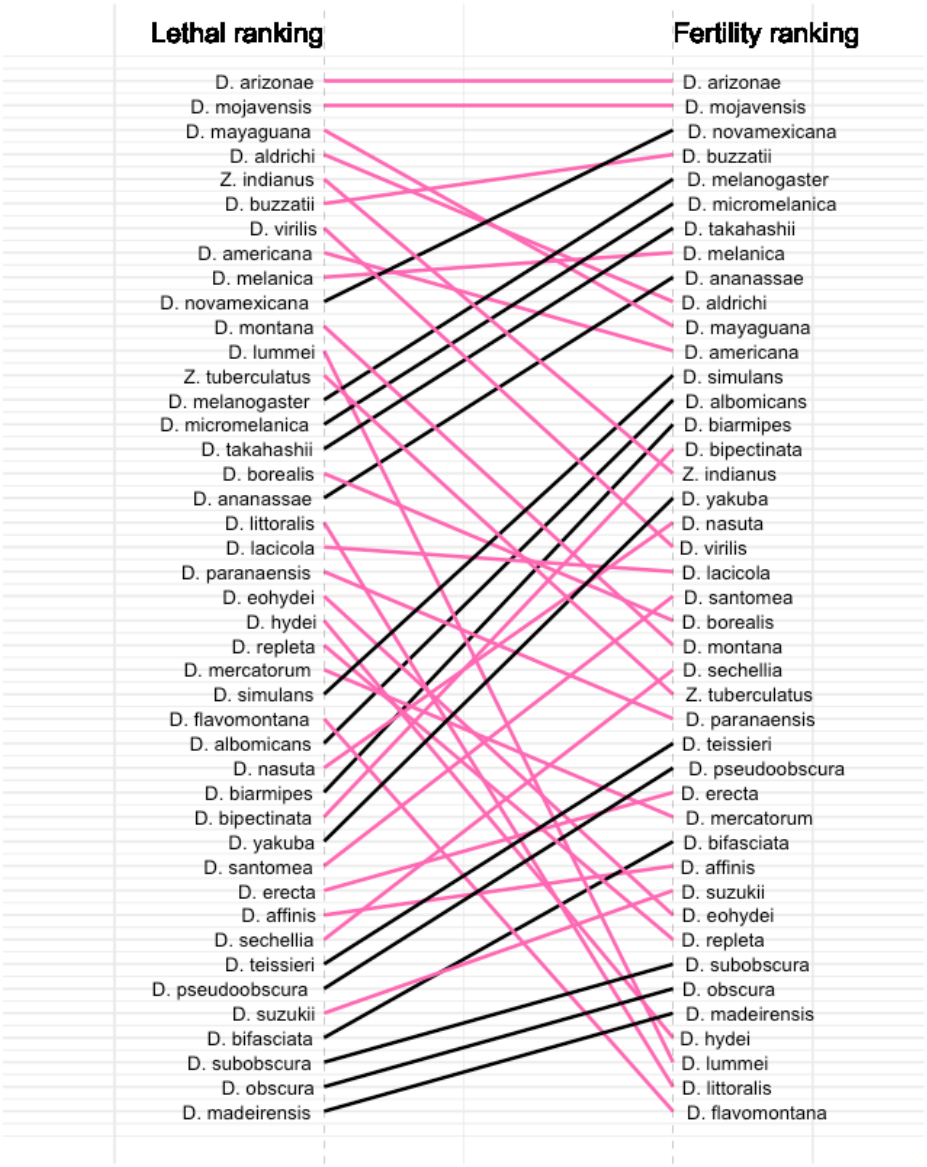


Fig S2: Relative ranking of species thermal robustness based on LT80 and TFL80. Most to least heat tolerant species ranked from top to bottom. Pink bars indicate species with significantly lower TFL than LT. Data based on TFL80 recorded 7-days post-heat shock.

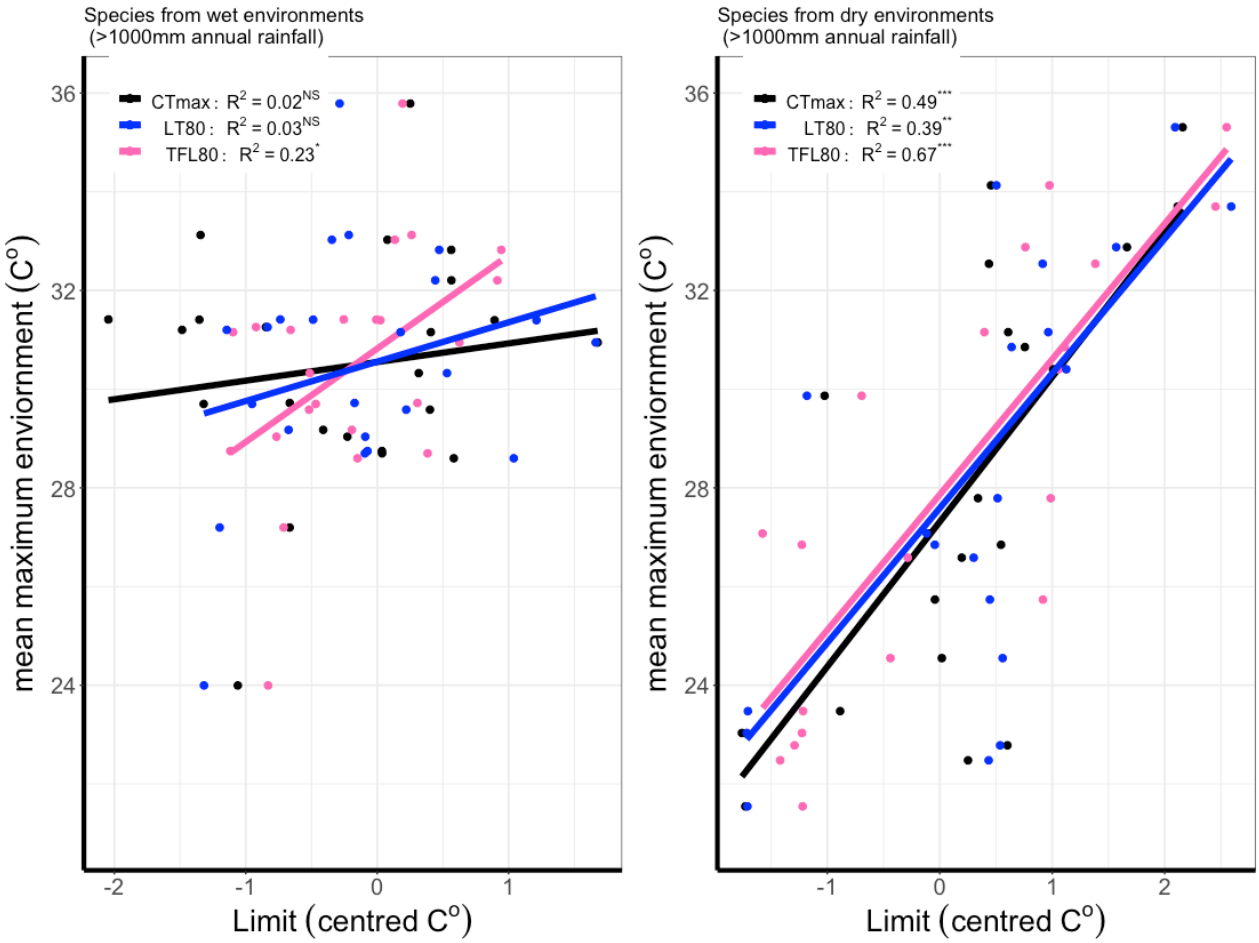


Fig S3. Correlations of maximum environmental temperature with CTmax, LT80 and TFL80. Upper thermal limits to fertility better predicts species' natural thermal habitat than either measure of critical limits; LT80 includes the same heat stress regime as TFL80, CTmax is widely used. To aid visualization here, data are split into 'wet' (>1000ml rainfall year^{-1}) and 'dry' (< 1000 ml rainfall year^{-1}) habitats as in(*4*). Model estimates from fitted lines here given in Table S2. Measures based solely on upper critical limits show no significant correlation for species that live in naturally wet habitat. This recapitulates the findings of Kellermann *et al*(*4*).

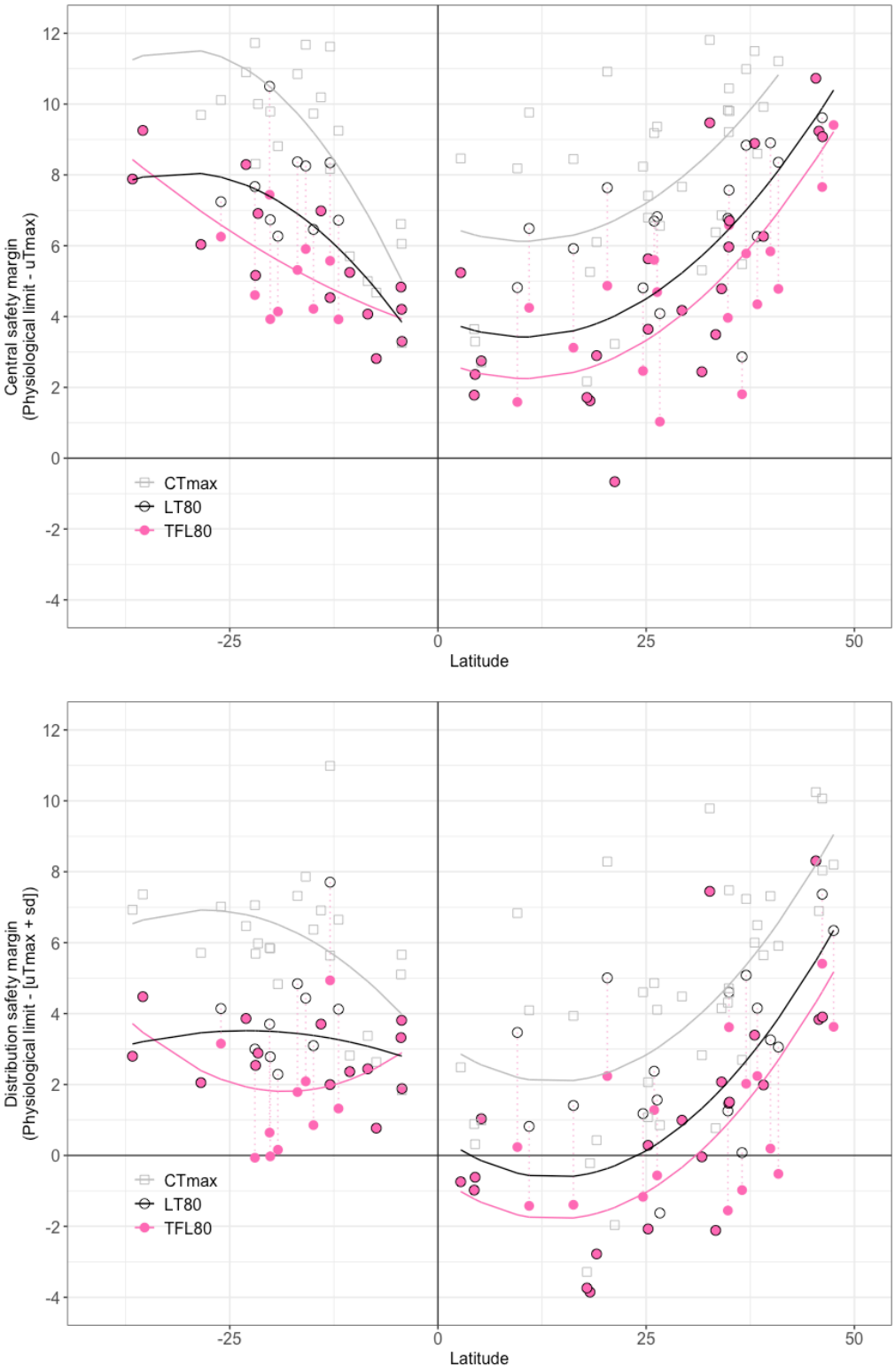


Fig S4: Thermal safety margins. Central (top) and Distribution (bottom) safety margins for species calculated using the physiological limit minus the mean maximum temperature across all recorded locations for each species. For distribution margins Tmax + 1sd is used to calculate safety margins, which has been argued captures populations at the boundary of a species' thermal range (*4*). Grey points and fit are safety margins calculated using knockdown CTmax

values. Black points and lines are safety margins based on LT80, and pink points are margins based on TFL80. Fitted lines are predictions from models in which the limit type (TFL, LT or CTmax) and latitude^2 are fixed effects and species identity is a random effect to account for repeated measures across species. Models and fits for North and South populations are calculated independently. Based on CTmax we predict that safety margins for central populations are 9.06 ± 3.24°C and populations at the thermal extremes have margins of 5.23 ± 3.24°C. This suggests that 3 species have populations at the edge of their range that are currently exposed to lethal temperatures in the hottest part of the year. LT80 is a more severe heat shock than CTmax and so inherently predicts smaller margins; 6.36 ± 2.92°C for central populations and 2.59 ± 3.08°C for populations at thermal extremes of species' ranges. This physiological measure predicts that 11 species have populations at the edge of their range that are currently exposed to lethal temperatures in the hottest part of the year. Including fertility data into LT80 to get TFL80 predicts that populations at the center of species' ranges have a 5.26 ± 2.55°C safety margin, whilst populations at the edge of species' thermal ranges have a 1.43 ± 2.8°C safety margin. TFL-based margins predict that 20 species have populations at the edge of their thermal range that are currently exposed to sterilizing temperatures in the hottest part of the year. Safety margins calculated with TFL80 are significantly smaller than those calculated from LT80 (Central safety margins: lmer: $F_1 = 28.04$, $P < 0.001$, estimate = -1.13, distribution safety margins; lmer: $F_1 = 17.33$, $P < 0.001$, estimate = -1.13). We did not statistically compare safety margins calculated from CTmax with those calculated from TFL80 and LT80 because CTmax inherently uses a different method of heating individuals.

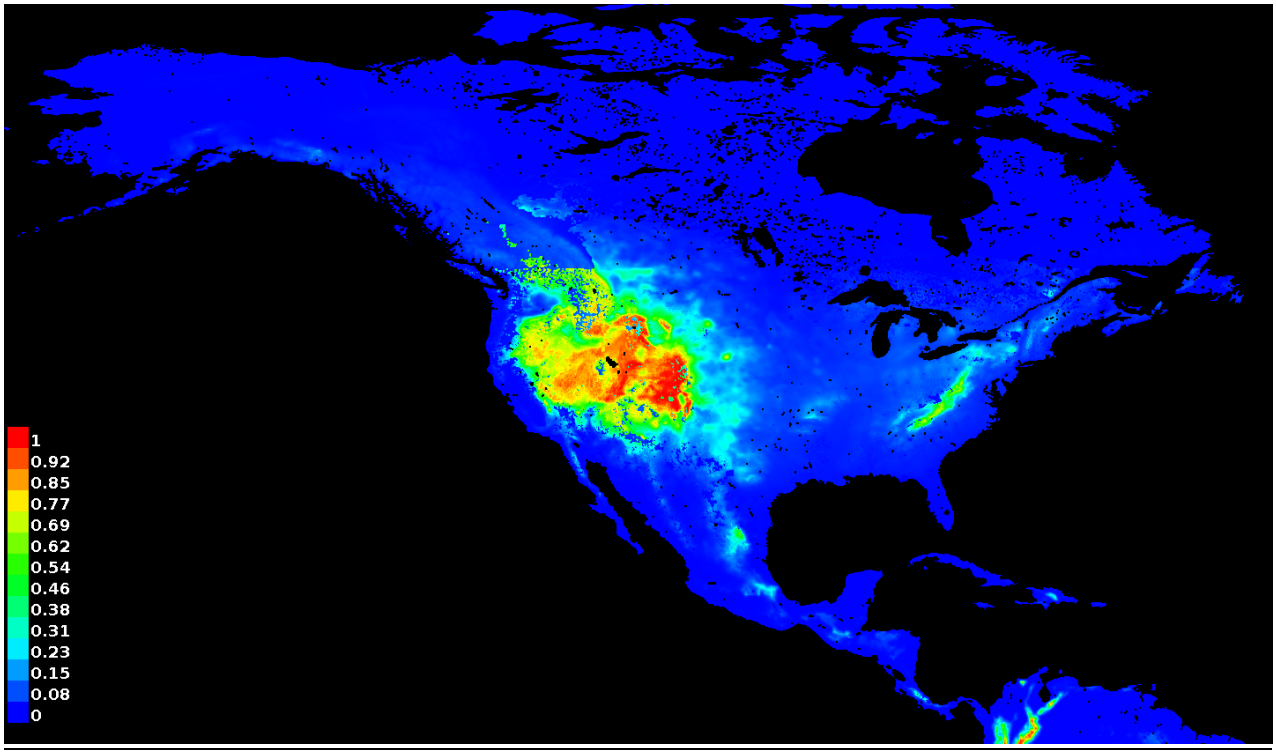


Fig S5: Base putative distribution model of *D. flavomontana*. Created with MaxEnt v3.4 with a 10-fold cross-validation run. The average test AUC for the replicate runs is 0.94, and the standard deviation is 0.058. We used the maximum training sensitivity plus specificity threshold to distinguish between suitable and unsuitable areas.

Table S1: Summaries of model that fit maximum environmental temperature (Tmax) to either CTmax, LT80 or TFL80 independently of each other. All models fit with upper physiological limits interacting with annual precipitation (Pann). R^2 given for maximal model. Terms retained after AIC model selection shown in italics. All models are phylogenetically controlled, with no significant phylogenetic signal found in any case (Pagel's lambda not significantly different to 0). All predictors centred and scaled.

Full Model	**Terms**	**DF**	**F**	**P**	**$^{adj}R^2$**
Tmax~CTmax*Pann	*CTmax*Pann*	*1,39*	*6.092*	*0.018*	*0.371*
	CTmax		9.351	0.004	
	Pann		12.282	0.001	
Tmax~LT80*Pann	LT80*Pann	1,39	0.995	0.325	0.306
	LT80		*14.311*	*<0.001*	
	Pann		*6.247*	*0.017*	
Tmax~TFL80*Pann	TFL80*Pann	1,39	0.531	0.471	0.544
	TFL80		*4.706*	*<0.001*	
	Pann		*3.551*	*0.001*	

628

Table S2: Summaries of model that fit either TFL80, LT80 or CTmax to Maximum Environmental temperature (Tmax). Species from 'wet' environments (Pann > 1000 ml) and 'dry' (Pann < 1000ml) modelled separately. All models are phylogenetically controlled, with no significant phylogenetic signal found in any case (Pagel's lambda not significantly different to 0).

Species group	Full Model	Terms	DF	F	P	$^{adj}R^2$
Dry	Tmax~CTmax	CTmax	1,18	19	<0.001	0.487
	Tmax~TFL80	TFL80	1,18	39.858	<0.001	0.672
	Tmax~LT80	LT80	1,18	13.47	0.002	0.396
Wet	Tmax~CTmax	CTmax	1,21	0.456	0.505	0.025
	Tmax~TFL80	TFL80	1,21	6.40	0.019	0.197
	Tmax~LT80	LT80	1,21	1.56	0.225	0.025

629

Table S3: Summary of model in which both TFL80 and CTmax are included as predictors of maximum environmental temperature (Tmax). R^2 shown for full model. Model was phylogenetically controlled, but no significant phylogenetic signal in residuals was detected. LT80 is excluded as a predictor based on high VIF scores indicating variance-inflating autocorrelation with both other predictors.

Full Model	Terms	DF	F	P	$^{adj}R^2$
Tmax~CTmax*Pann+TFL80*Pann	CTmax*Pann	1,37	0.164	0.688	0.531
	TFL80*Pann		0.329	0.569	
	CTmax		0.985	0.327	
	Pann		12.814	<0.001	
	TFL80		38.254	<0.001	

630

631

Table S4: Husbandry details for species used in this study. All species either *D* = *Drosophila* or *Z* = *Zaprionus*

Species	Rearing temp (°C)	StrainID & source	Multiple lines	Food	Mean age at heat (days)	Temperature stress range (°C)
D. affinis	18	DSSC Stock #: 14012-0141.14	N	M	7	29 – 36
D. albomicans	23	DSSC Stock #: 15112-1751.03	N	A	7	31 – 36
D. aldrichi	23	Wild caught single line, Texas 2019	N	A	7	33 – 40
D. americana	23	Single-line from (*35*)	N	M	7	33 – 38
D. ananassae	25	Single-line from (*35*)	N	A	7	30 – 38
D. arizonae	23	DSSC Stock #: 14012-0181.02	N	B	7	35 – 41
D. biarmipes	23	DSSC Stock #: 14023-0361.11	N	A	7	30 – 35
D. bifasciata	18	DSSC Stock #: 14012-0181.02	N	A	7	25 – 34
D. bipectinata	23	DSSC Stock #: 14024-0381.20	N	A	7	30 – 36
D. borealis	23	DSSC Stock #: 15010-0961.10	N	A	14	31 – 39
D. buzzatii	23	Single-line from (*35*)	N	M	7	32 – 39
D. eohydei	23	DSSC Stock #: 15085-1631.00	N	B	14	31 – 38
D. erecta	23	Single-line from (*35*)	N	A	7	29 – 39
D. flavomontana	23	Single-line from (*35*)	N	M	14	30 – 37
D. hydei	25	Mixed, long-term lab population at University of Liverpool	Y	B	14	29 – 37
D. lacicola	23	Single-line from (*35*)	N	M	7	31 – 37
D. littoralis	23	DSSC Stock #: 15010-1001.03	N	B	14	32 – 37
D. lummei	23	DSSC Stock #:	N	M	7	25 – 37

		15010-1011.07				
D. madeirensis	18	Mix of 3 lines, wild caught March 2019	Y	A	7	27 – 34
D. mayaguana	23	DSSC Stock #: 15081-1397.00	N	M	7	35 – 40
D. melanogaster	25	'Dahommey' lab population	Y	A	7	30 – 38
D. mercatorum	23	Wild-caught Madeira 2018, single line	N	A	7	31 – 40
D. micromelanica	23	DSSC Stock #: 15030-1151.01	N	A	7	30 – 38
D. melanica	23	DSSC Stock #: 15030-1141.03	N	A	7	32 – 38
D. mojavensis	23	Multi-line mixed population from (*35*)	Y	B	7	34 – 41
D. montana	18	Mixed population of IsoLines in (*36*)	Y	M*	21	31 – 37
D. nasuta	23	DSSC Stock #: 15112-1781.13	N	A	7	30 – 37
D. novamexicana	23	Multi-line mixed population from (*35*)	Y	B	7	28 – 38
D. obscura	23	Multi-line mixed population from (*35*)	Y	P	7	28 – 33
D. paranaensis	23	DSSC Stock #: 15082-1541.09	N	A	7	31 – 37
D. pseudoobscura	18	Mixed population if IsoLines in	Y	A*	7	29 – 35
D. repleta	23	DSSC Stock #: 15084-1611.01	N	A	7	30 – 38
D. santomea	23	Single-line rom (*35*)	N	A	7	29 – 37
D. sechellia	23	Multi-line mixed population from (*35*)	Y	P	7	29 – 37
D. simulans	25	Mixed of laboratory lines collected across Europe 2016 - 2018	Y	A	7	30 – 37
D. subobscura	18	Mixed population from Lines	Y	A	7	27 – 34

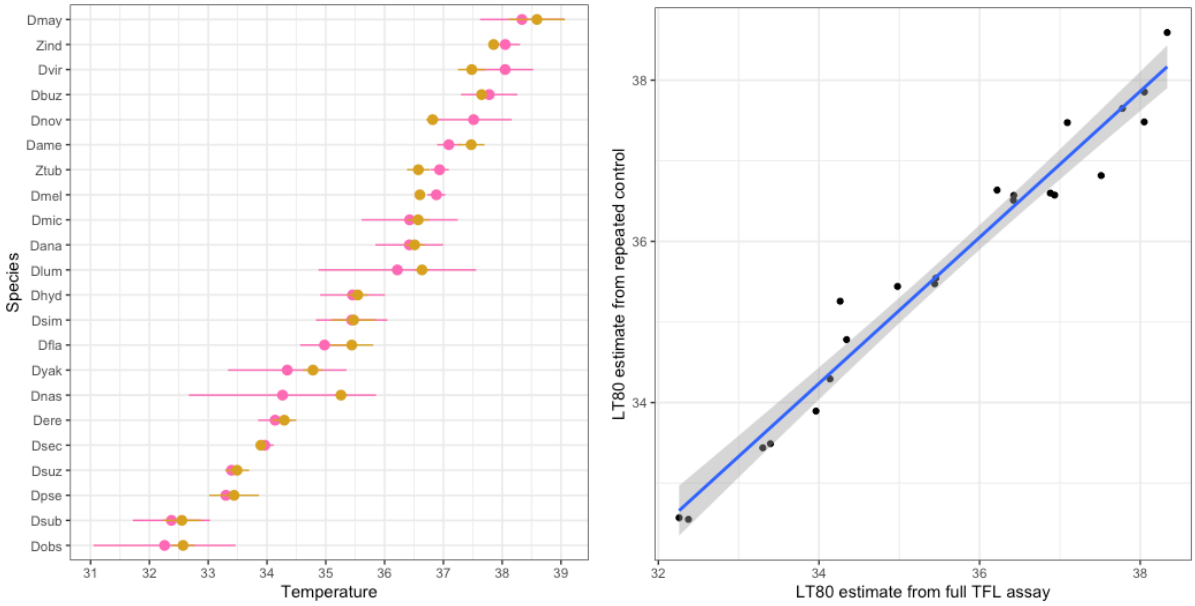
D. suzukii	23	Mixed populations of 3 IsoLines caught in Madeira (March 2018) and single lines each from the USA and Japan	Y	A	7	27 – 37
D. takahashii	23	Single-line from (*35*)	N	A	7	30 – 37
D. teissieri	23	Single-line from (*35*)	N	A	7	28 – 37
D. virilis	23	Multi-line mixed population from (*35*)	Y	P	7	26 – 38
D. yakuba	23	Multi-line mixed population from (*35*)	Y	A	7	3 – 38
Zaprionus indianus	23	Single line, wild caught 2019	N	B	7	33 – 39
Zaprionus tuberculatus	23	DSSC Stock #: 50001-0001.07	Y	B	7	33– 38

632

633 **Supplementary text**

634

635 *Repeatability of 4-hour heating assay:*



636

637 **Fig ST1.1: Repeatability of LT80**. Repeatability was high across the 22 *Drosophila* species

638 **tested**. Left panel: There was no significant difference in the estimate of LT80 between full

639 TFL runs and control repeats in any of these species (significance measured as non-

640 overlapping CIs of point estimates of LT80). Pink points = estimate from full TFL assay, gold

641 point = estimate from simultaneous multispecies LT80 assay. Right panel: The correlation

642 between the two independent measurements of LT80 across these species was strongly

643 positive and explained a high degree of variation in the data (coefficient = 1.06, R^2 = 0.96).

644

645

646

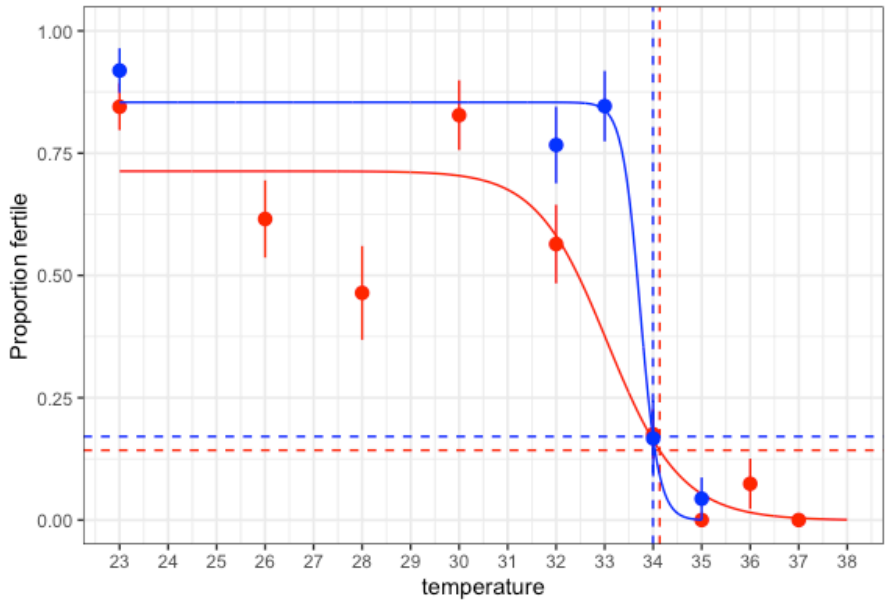
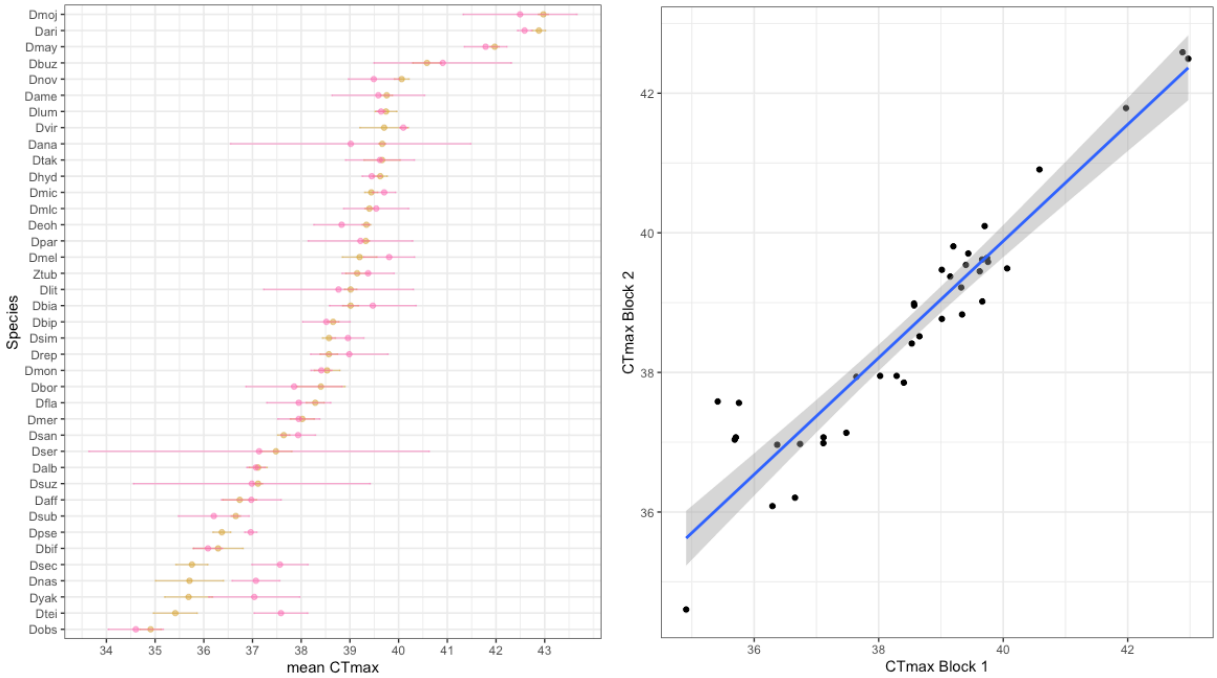


Figure ST1.2: Repeatability of TFL80 for *Drosophila virilis***.** We fully independently repeated the TFL80 assay for one species in which we found a significant difference between LT80 and TFL80. Two independent runs of sexually mature males were conducted by two researchers 6 months apart. Red = original data used to calculate TFL80 for *D. virilis* in this manuscript, blue = repeated assay. Dashed lines intersect at 80% threshold. Line fits predicted by 3-parameter dose-response model. The fly stock and equipment were identical.

673 *Repeatability of knockdown CTmax assay*

674

675



676

677

678 **Fig S2.1: Knockdown CTmax assays were generally highly repeatable.** Left panel; mean CTmax

679 temperature recorded from species' independent assays (gold points), and in mixed species

680 control blocks (pink points). There was no global significant difference between CTmax

681 estimates across all species (Block: $F_{1,831} = 3.246$, $P = 0.072$). However a significant experimental

682 block*species interaction was found for 4 of the 41 species tested (interaction term: $F_{40,831} =$

683 3.7589, P < 0,001); *Drosophila sechelia, D. nasuta, D. yakuba* & *D. tiesseri* all scored significantly

684 higher CTmax in our mixed species control block than in their own individual CTmax assays.

685 Right panel: The correlation between estimated CTmax values from both blocks was strongly

686 positive (coefficient = 1.07, $F_{1,37} = 309.84$, P < 0.001), and explained 89% of the variation.

687

688

689 *Correlations between TFL, LT and CTmax*

690 The correlation between TFL80 and LT80 is significant (PGLS: $F_{1,41} = 47.86$, $P < 0.001$, coefficient
691 = 1.631) but LT80 only explains just over half of the variation in TFL80 (PGLS: adj$R^2 = 0.53$,).
692
693 The correlation between absolute TFL80 and CTmax assays is significant (PGLS: $F_{1,41} = 19.40$, P
694 <0.001) but considerably weaker than a perfect proxy (PGLS: $R^2 = 0.30$, coefficient = 0.631).
695 These analyses were phylogenetically controlled, with no significant phylogenetic signal found
696 among model residuals ($\lambda = 0.543^{ns}$).
697
698 The correlation between the two measures of upper lethal limits (LT80 and knockdown CTmax)
699 is significant and positive (estimate = 0.668, $F_{1,41} = 89.176$, $P < 0.001$) but also shallower than
700 would be expected of exact proxies. However, this relationship explains 67.74% of the variation
701 in LT80 across species. LT80 and knockdown CTmax are different measures of upper thermal
702 limits and so deviation from parity is not unexpected. Notably, knockdown CTmax inherently
703 contains an element of thermal ramping, whilst LT80 employs a single static temperature.
704 These analyses were phylogenetically controlled, with a significant phylogenetic signal found
705 among model residuals ($\lambda = 0.570^{**}$).
706
707 The difference between TFL80 and LT80 (ΔLimits) shows no significant relationship with either
708 CTmax or LT80 (PGLS; knockdown CTmax: coefficient = 0.042, $F_{1,41} = 0.127$, $P = 0.72$, $R^2 = -0.02$;
709 LT80: coefficient = -0.015, $F_{1,41} = 0.011$, $P = 0.92$, $R^2 = 0.02$). This demonstrates that critical limits
710 cannot be used as a direct proxy for TFLs, and so TFLs are capturing biological variation that
711 measurements of critical limits do not. This also suggests that there is not simply an
712 unavoidable upper limit to fertility shared among species with differing upper critical limits.
713
714
715

716 *Phylogenetic signal in upper thermal limits*

717 Previous empirical work has found that absolute upper critical limits show strong phylogenetic
718 signal(*37*). This suggests that these limits may be difficult to rapidly evolve in response to
719 changes in the environment and suggests that species are sorted into their environment by
720 their evolutionary history, rather than locally adapted to it(*4*). We find similar degrees of
721 phylogenetic signal in our measures of critical upper limits (LT80 $\lambda = 0.94$ $P_0 < 0.001$, $P_1 = 0.108$;
722 $\kappa = 0.866$, $P = 0.001$), (CT_{max} $\lambda = 0.853$, $P_0 < 0.001$, $P_1 < 0.001$; $\kappa = 0.51$, $P < 0.001$). We also find
723 phylogenetic signal in our estimate of TFL80, however it is less strong than the signal for either
724 upper lethal limit ($\lambda = 0.722$, $P_0 = 0.01$, $P_1 = 0.004$; $\kappa = 0.32$, $P = 0.005$). The difference between
725 LT80 and TFL80 (ΔLimits hereafter) are explained by evolutionary relatedness between species
726 ($\lambda = 0.823$, $P_0 = < 0.001$, $P_1 = 0.024$, $\kappa = 0.857$, $P = 0.001$), but this is primarily because the
727 majority of species in the *Sophophora* subgenus have a ΔLimit of 0°C. Whether this means that
728 species are able to rapidly evolve TFLs within the range of their ΔLimits is an open question.