Journal of Evolutionary Biology, 2024, **XX**, 1–15
https://doi.org/10.1093/jeb/voae021
Advance access publication 13 February 2024
Methods Article

JOURNAL OF
Evolutionary Biology
eseb

Systematic approaches to assessing high-temperature limits to fertility in animals

Amanda Bretman[1], Claudia Fricke[2], Julian Baur[3], David Berger[3], Merel C. Breedveld[4], Diego Dierick[5], Berta Canal Domenech[6], Szymon M. Drobniak[7,8], Jacintha Ellers[9], Sinead English[10], Clelia Gasparini[4], Graziella Iossa[11], Malgorzata Lagisz[8], Shinichi Nakagawa[8], Daniel W.A. Noble[12], Patrice Pottier[8], Steven A. Ramm[13], Melissah Rowe[14], Eva Schultner[15], Mads Schou[16], Pedro Simões[17,18], Paula Stockley[19], Ramakrishnan Vasudeva[1], Hester Weaving[10], Tom A.R. Price[20], Rhonda R. Snook[21]

[1]Faculty of Biological Sciences, University of Leeds, Leeds, LS2 9JT, United Kingdom
[2]Institute for Zoology, Halle-Wittenberg University, Halle (Saale), Germany
[3]Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden
[4]Department of Biology, University of Padova, Padova, Italy
[5]La Selva Biological Station, Organization for Tropical Studies, San Pedro Montes de Oca, Costa Rica
[6]Institute for Evolution and Biodiversity, University of Muenster, Muenster, Germany
[7]Institute of Environmental Sciences, Jagiellonian University, Kraków, Poland
[8]School of Biological, Earth and Environmental Sciences, The University of New South Wales, Sydney, NSW, Australia
[9]Amsterdam Institute for Life and Environment, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[10]School of Biological Sciences, University of Bristol, Bristol, United Kingdom
[11]School of Life and Environmental Sciences, University of Lincoln, Lincoln, United Kingdom
[12]Division of Ecology and Evolution, Research School of Biology, The Australian National University, Canberra, ACT, Australia
[13]UMR 6553 Ecobio - Ecosystèmes, Biodiversité, Evolution, Université de Rennes, Rennes, France
[14]Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands
[15]Zoology/Evolutionary Biology, University of Regensburg, Regensburg, Germany
[16]Department of Biology, Aarhus University, Aarhus, Denmark
[17]cE3c—Centre for Ecology, Evolution and Environmental Changes & CHANGE—Global Change and Sustainability Institute, Lisboa, Portugal
[18]Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal
[19]Mammalian Behaviour & Evolution Group, Department of Evolution, Ecology and Behaviour, University of Liverpool, Neston, United Kingdom
[20]Department of Evolution, Ecology and Behavior, University of Liverpool, Liverpool, United Kingdom
[21]Department of Zoology, Stockholm University, Stockholm, Sweden

Corresponding author: Amanda Bretman, Faculty of Biological Sciences, University of Leeds, Clarendon Road, Leeds LS2 9JT, United Kingdom. Email: a.j.bretman@leeds.ac.uk

Abstract

Critical thermal limits (CTLs) gauge the physiological impact of temperature on survival or critical biological function, aiding predictions of species range shifts and climatic resilience. Two recent *Drosophila* species studies, using similar approaches to determine temperatures that induce sterility (thermal fertility limits [TFLs]), reveal that TFLs are often lower than CTLs and that TFLs better predict both current species distributions and extinction probability. Moreover, many studies show fertility is more sensitive at less extreme temperatures than survival (thermal sensitivity of fertility [TSF]). These results present a more pessimistic outlook on the consequences of climate change. However, unlike CTLs, TFL data are limited to *Drosophila*, and variability in TSF methods poses challenges in predicting species responses to increasing temperature. To address these data and methodological gaps, we propose 3 standardized approaches for assessing thermal impacts on fertility. We focus on adult obligate sexual terrestrial invertebrates but also provide modifications for other animal groups and life-history stages. We first outline a "gold-standard" protocol for determining TFLs, focussing on the effects of short-term heat shocks and simulating more frequent extreme heat events predicted by climate models. As this approach may be difficult to apply to some organisms, we then provide a standardized TSF protocol. Finally, we provide a framework to quantify fertility loss in response to extreme heat events in nature, given the limitations in laboratory approaches. Applying these standardized approaches across many taxa, similar to CTLs, will allow robust tests of the impact of fertility loss on species responses to increasing temperatures.

Keywords: reproduction, heat, thermal fertility limit, thermal sensitivity of fertility, critical thermal limit

Glossary: CTL: Criticsal thermal limit, the upper or lower temperature at which critical biological function (often measured as motor control or coordinated movement) is lost, or death occurs; CT_{max}: Critical thermal maximum, the highest temperature at which a physiological function is lost (often measured as motor control or coordinated movement), or death occurs; LT_{80}/LT_{50}: Lethal temperature, temperature at which there is either 80% or 50% mortality in a population/ set of experimental organisms; TFL: Thermal fertility limit, temperature at which individuals become (at least temporarily) sterile (scored as a binary 0/1 outcome);

Received October 6, 2023; revised January 23, 2024; accepted February 7, 2024

TFL_{80}/TFL_{50}: Temperature at which either 80% or 50% of individuals in a population/ set of experimental organisms are sterile; TSF: Thermal sensitivity of fertility, the relationship between temperature and reproductive ability. Here the number of offspring produced or proxies thereof are a measure of fertility rather than the binary 0/1 score of fertility

Introduction

Background

Current and looming climatic changes caused by recent anthropogenic activities have generated catastrophic consequences for biodiversity (Pecl et al., 2017). Climate models predict both continued warming and more frequent, intense, and longer-lasting heatwave events (Calvin et al., 2023). Therefore, understanding how changes in temperature regimes will affect species' ranges and population viability is vital to improving forecasts of future biodiversity patterns (Urban et al., 2016). To date, studies of organismal thermal biology have typically focussed on how critical thermal and lethal temperatures constrain species ranges (Deutsch et al., 2008; Kellermann et al., 2012; Pinsky et al., 2019; Sunday et al., 2012). This focus has employed a critical thermal limits (CTLs) framework (Lutterschmidt & Hutchison, 1997; Terblanche et al., 2007) aimed at defining the range of temperatures in which an organism can function. The high-temperature endpoint of thermal tolerance is frequently defined as the upper thermal limit (CT_{max}), typically measured in animals as the temperature at which individuals lose either motor coordination or physiological function. The CTL framework has been widely used across different taxa, and the accumulation of these data provides a powerful tool to help understand how temperature impacts broad physiological and ecological properties of species, for example, geographical distribution (Bennett et al., 2021; Bush et al., 2016; Kellermann et al., 2012; Pinsky et al., 2019; Sunday et al., 2012). CTLs can be measured following different approaches in which organisms are exposed to either static (Hoffmann et al., 2002) or ramping (Terblanche et al., 2007) temperatures to define a species' thermal tolerance. However, the initial lack of standardization in CT_{max} assays resulted in methodological issues that questioned its relevance in estimating species' upper thermal limits (Leong et al., 2022). Additionally, a recent meta-analysis found that the chosen trait measured to determine CTL (e.g., loss of activity, adhesion to a surface, or death) used in ramping assays significantly affected observed plasticity in CTL (Weaving et al., 2022), highlighting that choosing a consistent approach can yield more comparable data. Moreover, while powerful, CTLs do not always correlate well with current species distributions (Gouveia et al., 2014; Sunday et al., 2012), potentially because thermally sensitive traits not traditionally measured in the CTL framework could contribute to limiting species distributions and population persistence (Walsh, Parratt, Hoffmann, et al., 2019).

One such trait is fertility. The negative effects of high temperatures on fertility are well known (Walsh, Parratt, Hoffmann, et al., 2019). Given that reproduction is essential for population persistence, if fertility is more thermally sensitive than survival traits, then this may have a profound effect on population viability and how species respond to climate change. Two recent studies across nearly 50 *Drosophila* species have taken an approach similar to the static CTL method to determine a species' thermal fertility limit (TFL; Parratt et al. 2021; van Heerwaarden & Sgrò, 2021). TFLs consider fertility as a binary trait (1 = fertile, producing at least one offspring; 0 = sterile, no offspring) and identify the temperature at which a proportion of individuals in a population become (at least temporarily) sterile (e.g., TFL_{50} is the temperature at which 50% of individuals are sterile, TFL_{80}, when 80% are sterile). These studies revealed that TFLs can occur at lower temperatures than critical and/or lethal limits (Parratt et al., 2021; van Heerwaarden & Sgrò, 2021) and that these survival measures are often poor predictors of TFLs (Parratt et al., 2021). For example, while TFL_{80} was on average 1.15 °C lower than LT_{80} (the lethal limit, which is the temperature at which 80% of individuals die) in 43 *Drosophila* species, the difference between the two values ranged from 0 to 4.3 °C (Parratt et al., 2021). Thus, TFLs cannot readily be predicted from survival temperatures, and consequently, their relationship to other critical limits must be experimentally determined. Moreover, TFLs were better predictors of *Drosophila* species distributions, with TFLs predicting both more restricted range size under projected climate change scenarios and increased risk of extinction under laboratory conditions relative to CT_{max} (Parratt et al., 2021; van Heerwaarden & Sgrò, 2021).

These two pioneering studies in *Drosophila* strongly suggest that species distributions might be limited by the temperature at which reproduction becomes impossible and that thermally induced sterility can result in population extinction faster than temperatures, causing physiological failure and death. If fertility is lost at lower than lethal temperatures in a wide range of organisms, then current predictions based on survival limits may have overestimated the resilience of many organisms to climate change. Thermal consequences for fertility, therefore, have both fundamental relevance to improving predictions on species vulnerability and future distributions and applied relevance to conservation and food security (i.e., reproductive capabilities of livestock and pollinators). We therefore argue that researchers should begin quantifying both survival and fertility limits. These data will be imperative for determining the extent to which these limits differ across taxa and how such differences alter and improve forecasts of species responses to warming globally.

Here, we propose standardized methodologies to collect robust data for this goal. While many studies have reported fertility consequences in response to heat stress (see Dougherty et al., 2024; Walsh, Parratt, Hoffmann, et al., 2019 for reviews), these data have not been collected in a standardized way that can maximize their use to evaluate responses to climate change (Dougherty et al., 2024). For example, TSF-type designs differ in the duration and intensity of stress, the life stage tested, whether the stress is fluctuating or constant, and whether there was a period of acclimation prior to the stress. Additionally, only the two studies on *Drosophila* have employed a TFL approach. As we anticipate studies of thermal effects on reproduction will be an increasingly important research area, we set out three protocols to guide this future effort and avoid methodological pitfalls. We focus our protocols on obligate sexual terrestrial invertebrates, which represent important ecological groups and some of the most speciose taxa (Eisenhauer et al., 2019). Moreover, these taxa are relatively straightforward to assay, therefore paving the way for large-scale comparative studies. However, such protocols should also be widely applicable to other animal groups, so we provide some taxon-specific modifications in appendices.

Table 1 Data information checklist. This checklist is intended to ensure researchers are collecting standardized data, with a view to incorporation in a future database. It is not intended to be an exhaustive description of the experimental design, and authors will include much more information in a publication. We also expect that authors, as a standard, make raw data openly available with publications.

Animal descriptors	
Taxonomic information	Phylum, Class, Order, Family, Genus, Species as per NCBI
Habitat	Terrestrial, aquatic or both
Fertilization mode	Internal, external or sperm caster
Latitude/Longitude location	Country or sea of origin and latitude/longitude or GPS coordinates
Elevation	Source population height above sea level
Rearing condition	Laboratory culture, wild-caught, farmed, or managed in natural conditions
Generations in lab	If applicable, the number of generations kept in laboratory conditions prior to the experiment
Number of founders	If applicable, the number of individuals from which the laboratory/managed culture originated
Prior conditions	For laboratory populations, standard conditions such as temperature, light regime, humidity, mating regime (e.g., panmixis)
Experiment descriptors	
Experiment type	Experimental manipulation of temperature in the laboratory or after a naturally high-temperature event for wild-caught in the lab or measured in the wild
Sex experiencing thermal manipulation	Male, female, both sexes, hermaphrodite, asexual (or clone)
Age of fertility measurement	Age of individuals when fertility is measured
Age of heat exposure	Age, or life stage if using the developmental protocol, at which the heat stress is applied
N	The final total number of individuals in each experimental treatment group (not a range)
Rearing/benign temperature	The standard (typically benign) rearing temperature of the laboratory population If using wild-caught animals, the mean daily maximum for 10 days prior to the heat event
Experimental temperatures	Temperatures at which fertility is measured in the laboratory. If using wild-caught animals, the maximum temperature is reached during the heat event.
Lab acclimation	If wild-caught animals were acclimated to laboratory conditions before assay, how long for and under which conditions (humidity, rearing regime, temperature)
Heat duration	Following the laboratory-based TFL or TSF protocol, this will be 4 hr If using wild-caught animals, how long they were exposed to the higher temperatures of the natural heat event
Fertility metric	Number of individuals producing live offspring, live sperm/eggs Number of offspring, sperm count, sperm viability, sperm morphology or sperm swimming speed, egg count or egg viability, gamete DNA damage, gonad deformation/size
When measured	Number of days over which fertility was measured after heat shock. This may be more than one time point
TFL specific	
LT_{80}	°C temperature of 80% death for each group, e.g., sex or age
TFL_{80}	°C temperature of sterility in 80% for each group, e.g., sex or age
LT_{50}	°C temperature of 50% death for each group, e.g., sex or age
TFL_{50}	°C temperature of sterility in 50% for each group, e.g., sex or age
TFL recovery	If measured at multiple time points, the proportion of individuals showing recovery
CT_{max}	The upper thermal limit in °C
CT_{max_source}	This study or gives the reference DOI
TSF specific	
TSF	Fertility metric for each temperature/sex/time point measured. Even if data are non-normally distributed, providing a standardized effect size and/or mean and SD is helpful for meta-analyses
TSF_{max}	Maximum reduction in the trait value as % of performance under benign laboratory maintenance temperature or pre-heatwave measurement
$TSF_{recovery}$	If measured sometime after the heat event/lab manipulation, % increase in fertility from TSF_{max}
%_recovery	If measured sometime after the heat event/lab manipulation, % recovery in fertility compared to pre-stress

We first outline a "gold standard" approach in which survival and TFLs are simultaneously determined after short periods of heat stress, similar to the static CTL approach. The rationale for this is that anthropogenic atmospheric inputs are altering global thermal regimes, increasing mean temperature, and causing more frequent, intense and prolonged extreme heat events, i.e., heatwaves (Calvin et al., 2023). Periods of extreme temperatures may have more important impacts on the persistence and adaptation of natural populations than gradual increases in global mean temperature (Murali et al., 2023; Stillman, 2019). While this is our focus, we acknowledge that various thermal conditions may be relevant for understanding thermal effects on fertility. We then outline an alternative protocol in which the thermal sensitivity of fertility (TSF; see Glossary) is determined by comparing reproductive output after exposure to only two temperatures (benign and stressful). Finally, we provide a standardized protocol for assessing the real-world fertility implications during naturally occurring heat waves. We provide guidance on determining the temperatures to consider, sample size and analytical framework for these protocols. To facilitate future comparative work, modelling approaches and/or meta-analyses, we provide a checklist for metadata that should be collected when designing these experiments (Table 1). While not exhaustive, we hope researchers working on other taxa (e.g., plants, microbes, fungi) will use these recommendations and the spirit of standardization to develop protocols amenable to their taxonomic group.

Our goal is to unify the rapidly expanding field of research on the effects of temperature on fertility by proposing pragmatic and standardized approaches. We acknowledge that these are a starting point for a full investigation of high-temperature impacts on fertility, for example, because we focus on acute stress. However, we argue these data are fundamental to building an urgently needed and robust knowledge base to predict how increasing temperatures impact biodiversity. There are a variety of open fundamental questions about the impact of temperature on fertility that currently cannot be resolved due to a lack of comparable data (Iossa, 2019; Walsh, Parratt, Atkinson, et al., 2019; Walsh, Parratt, Hoffmann, et al. 2019). The framework we suggest here can ultimately contribute to addressing such questions, including (a) Are TFLs predominantly below lethal thermal limits? (b) Is male or female fertility more vulnerable? (c) Are ectotherms more vulnerable than endotherms? (d) Do different life stages differ in their vulnerability to heat stress? (e) To what extent does knowledge of these various limits and sex- and taxa-specific responses improve the ability to forecast species responses to future climate scenarios? Answers to these questions will provide the necessary refinements to better predict species' vulnerability to increasingly high temperatures as a consequence of climate change.

Aims and remit of our standardized approaches

Our focus here is to develop a straightforward, unified approach to assess the limits to fertility after short periods of heat stress, as might be experienced in the hottest part of the day during extreme heat events. Moreover, this mimics CTL experiments, of which many are based on a brief heat shock. Other protocols would be necessary to evaluate the impact of chronic heat stress. Below we outline our protocols for obligate sexual terrestrial invertebrates with a focus on internal fertilization. The TFL approach is based on that of Parratt et al. (2021), while the TSF approach is based on that of Baur et al. (2022). We also provide accompanying appendices where we outline how protocols can be adapted for other animal taxa (Supplementary Material 2.1 reptiles, Supplementary Material 2.2 fish, Supplementary Material 2.3 birds, Supplementary Material 2.4 mammals).

In the main text, we focus on measuring the consequences for fertility when mature adults are exposed to heat stress, as it is easier to measure fertility consequences in real-time in reproductively active individuals. Nevertheless, short exposure to extreme heat stress at earlier life stages might negatively impact subsequent reproduction, and some developmental stages might be even more sensitive than adults (Sales et al., 2021; Vasudeva et al., 2021). For this reason, we provide a modified protocol (Supplementary Material 1.4) for assessing how heat stress during development may impact adult fertility.

To accurately define TFLs, we recommend highly controlled experiments, typically performed in a laboratory setting (see *Protocol for determining thermal fertility limits* section). Such experiments provide a basis for comparison across taxa. However, not all taxa are amenable to such experimental approaches. Moreover, laboratory-based experiments cannot fully capture the natural thermal variation organisms experience that may either elevate or reduce temperature effects on fertility (e.g., cumulative negative effects or hardening, respectively). We, therefore, envision a "natural history of thermal sensitivity of fertility" approach both to complement laboratory-based estimates of the effects of extreme heat on fertility and to directly estimate the effects of extreme heat events on fertility in wild populations. Previous estimates of thermal effects on fertility in wild populations have been based on long-term monitoring, frequently of managed populations, linking variation in fertility and weather data (Peña et al., 2019; Schou et al., 2021). With the pressing need to understand the effects of increasing temperatures on population persistence and the relative paucity of long-term population monitoring for fertility, we suggest an alternative approach that exploits naturally occurring high-temperature episodes (see *A natural history of thermal sensitivity of fertility* section). Laboratory-based data harnesses the full power of the TFL approach, providing high resolution in determining these limits. Combining these insights with data from the field will be extremely powerful as this will (a) allow access to taxa that are not amenable to laboratory experiments, (b) give real-time insights into thermal fertility dynamics in nature, and (c) allow measurements under co-varying climate factors such as radiation and humidity (which typically are controlled in laboratory-based measurements). However, identifying TFLs in natural populations is difficult given that thermal fertility sensitivity across a wide range of temperatures (i.e., thermal reaction norms) usually cannot be determined. Furthermore, it may be impossible to measure fertility in some wild-caught species if they are not amenable to producing offspring in the lab. Instead, reporting proxies of the sensitivity of fertility to heat stress may be useful (i.e., TSF, *Protocol for thermal sensitivity of fertility* section).

We encourage open and responsible research, for example, outlining approaches to data archiving in *Data recording and accessibility* section. If collecting organisms from the wild, whether to establish lab stocks or to measure TSF from natural populations, then collection permits may be necessary. Researchers may also need to consider the movement of

genetic resources across borders under the Nagoya protocol (https://www.cbd.int/abs). When collecting in countries to which the researcher is not local, particularly Global North researchers collecting in the Global South, we advocate avoiding "helicopter research" and fostering collaborative science (Haelewaters et al., 2021). Collections in the wild should not negatively impact effective population size, and such decisions should include recognizing that heat stress may have a negative impact on fertility, further reducing population size.

Protocol for determining thermal fertility limits

Here, we describe an ideal standardized laboratory methodology that requires the availability of large numbers of individuals, groups of which can be exposed to several highly controlled temperatures in parallel, and in which critical life-history factors such as mating history, age, sex, and previous thermal experience can be standardized. Prior knowledge of the species being assayed is required, such as the age of reproductive maturity, and elements of the protocol can be adjusted to the species used. If species-specific considerations necessitate a change in the protocol, then these should be carefully and explicitly reported. An overview of the protocol is given in Figure 1 and the steps outlined in Figure 2.

Equipment for thermal manipulation

We intend the protocol to be widely usable, but its implementation will vary depending on the availability of equipment. Heat may be manipulated with climate chambers, warming plates, or water baths. It is worth considering whether this introduces pseudoreplication, as depending on the availability of equipment, animals kept in the same chamber or controlled temperature room may be more likely to show more similarity in performance. However, we think this is unlikely to have major effects because of the short duration of heat stress we recommend (*Simultaneous measurement of lethal limits and thermal fertility limits* section).

To choose the temperatures assayed, the margin of error of the chosen equipment must be considered. For example, if an incubator has an error of ±0.5 °C, then the minimum difference between assayed temperatures should be 1.5 °C. This is because the set temperature of the lower temperature incubator (e.g., 25 °C) could be running 0.5 °C higher, and the set temperature of the higher temperature incubator (e.g., 26 °C) could be running 0.5 °C lower so that for much of the heat stress time the incubators do not actually differ in temperature. For small terrestrial invertebrates, the use of a water bath is likely the most accurate as it can maintain specific temperatures for long periods. Irrespective of the equipment, temperature should be monitored during heat stress using temperature probes, taking care to assess the temperatures the animals are experiencing. Humidity may contribute to variation in CTL (e.g., Riddell et al., 2023), but likely this only becomes problematic for assays conducted for over 10 hr, even in small insects (Terblanche et al., 2011), whereas we advocate a substantially shorter assay (Section 2.3). Nevertheless, to avoid any confounds of desiccation, we recommend assay tubes contain an amount of standard food or damp cotton wool, though note that including these may reduce the realized temperature within the tube (Terblanche et al., 2011). If possible, humidity should be monitored in the most relevant way (i.e., if the animals are in food vials, then the humidity within the vial is the most relevant measure, not the humidity in the incubator).

Experimental animals

We recommend the use of sexually mature experimental individuals raised under standardized conditions, drawn from a laboratory population that has been maintained for multiple generations under these same conditions. If the population has been recently established from the wild using recently caught stocks, it could mean the population is under stress from adaptation to the laboratory. To avoid this and any transgenerational/parental non-genetic effects, we recommend at least two (non-overlapping) generations of laboratory acclimation before use in experiments (Hoffmann & Ross, 2018). However, long-term maintenance could result in laboratory adaptation to different selection pressures than those experienced in nature (Hoffmann & Ross, 2018). Both issues may confound laboratory-based fertility and survival measures to thermal stress. However, no evidence of laboratory adaptation confounding TFL measurements was found in 43 *Drosophila* species (Parratt et al., 2021). Moreover, the somewhat simple laboratory measures were able to successfully predict current natural distributions. Furthermore, a meta-analysis on insects found no differences in CTL plasticity between laboratory and field populations (Weaving et al., 2022). These results suggest thermal limit responses in laboratory populations broadly reflect those in the wild, though the extent to which extrapolation from lab to field is possible may be taxon-specific (Morgan et al., 2019). Regardless, the source stock (i.e., location collected), number of generations in the laboratory, an estimate of population size (notes on bottlenecks or number of founders where possible) and standard rearing conditions should be reported when known. A further potential consideration is whether the animals are carrying a pathogen, as infection can alter thermal tolerance (Hector et al., 2023; Porras et al., 2020).

One major unresolved question is whether the fertility of opposite sexes—or individuals exhibiting different sex roles in the case of hermaphrodites (see Supplementary Material 1.2)—is differentially susceptible to heat stress (Iossa, 2019). Thus, the gold standard is to measure both male and female fertility independently. Since the TFL is measured as the proportion of a population or cohort of test animals that is sterile at a given temperature, assessment occurs by mating the focal thermally-treated sex to control temperature-treated individuals of the opposite sex. To achieve this, a plentiful supply of both sexes must be available (see *Sample size* section), and the experiment needs to be conducted with both males and females as focal individuals separately (see *Simultaneous measurement of lethal limits and thermal fertility limits* section). Subjects will, therefore, need to be maintained until maturity, likely in single-sex groups, to ensure they remain virgin for many species (e.g., vertebrates will require different approaches as outlined in Supplementary Materials 2.1–2.4). If contact with the opposite sex is necessary to reach sexual maturity, then a period of isolation before the assay should be implemented to ensure full sperm stores in males and, ideally empty sperm stores in females. For standardization and comparison across taxa, the TFL should be determined after reaching sexual maturity in healthy individuals.

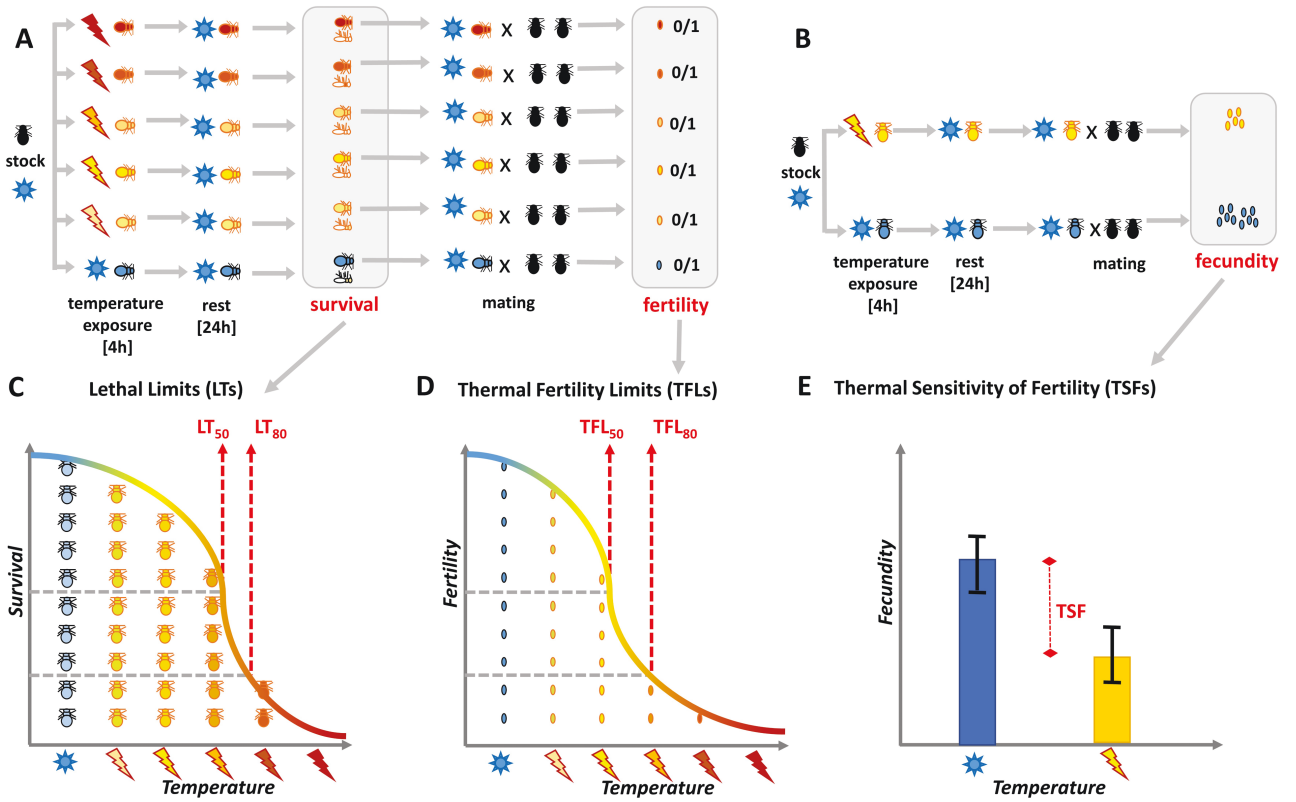

Figure 1. An overview of the gold-standard laboratory protocols (A) to simultaneously assay Lethal Limits and Thermal Fertility Limits or (B) Thermal Sensitivity of Fertility. Age-matched adults from a stock population (black beetle symbols) are raised at the benign temperature (sun symbols). When sexually mature, experimental individuals (beetle symbols other than black) are then exposed for 4 hr to the benign temperature or varying higher temperatures (lightning bolt symbols). 24 hr after temperature exposure (C) LT is then measured via the number alive/dead. Survivors are mated and scored as fertile/sterile (D) to determine TFL, though number of offspring from those that are fertile can also be counted. An additional measurement of TFL should be made at a later time point. Additionally (B), the Thermal Sensitivity of Fertility (TSF) can be measured as the change in number of offspring (E) at benign versus a higher temperatures (e.g., the ratio between temperatures). TSF can be used as an alternative if there are limitations to following the gold-standard protocol or when collecting real-time data from animals experiencing stressful temperatures in the wild.

Thus far, we have assumed working with obligate sexually reproducing terrestrial invertebrates in which the male places his sperm directly in or on the female genital opening. Other modes of sperm transfer involve the external deposition of sperm by the male, often packaged into a protective spermatophore. This spermatophore may be waiting several days to be picked up by a female. Deposition of the spermatophore on a substrate in the environment increases exposure to external conditions and hence constitutes an additional phase in which damage due to heat stress may occur. Completely dissociated sperm transfer is rather common among terrestrial invertebrates, including several species of millipedes, pseudoscorpions, mites and springtails (Proctor, 1998; Witte & Döring, 1999). For species with dissociated sperm transfer, an additional exposure treatment that tests the TFL of deposited spermatophores is necessary. To achieve this, males kept under benign (control) conditions should be allowed to deposit spermatophores, which can then be exposed to temperatures as described for the TFL protocol. After the heat exposure, the vials with spermatophores should be returned to their maintenance temperature and offered to virgin, receptive females.

Simultaneous measurement of lethal limits and thermal fertility limits

To assess whether fertility limits occur at lower temperatures than lethal limits, we outline a protocol that measures both simultaneously by exposing individuals to static temperature manipulations. For both measures, limits are defined as when 80% of individuals tested die (LT_{80}) and when 80% of living individuals are sterile (TFL_{80}). We score individuals as fertile if they produce at least one live offspring; otherwise, they are scored as sterile. Sterility then includes reductions driven by damage to physiological and behavioural aspects of reproduction. Thus, for both LT and TFL, an individual's outcome is binary: individuals are either alive or dead, and if alive, they are either fertile or sterile.

A thermal reaction norm approach can be used to determine both LT_{80} and TFL_{80} simultaneously (Figures 1 and 2) to assess the temperatures when the respective limits are reached. Because identifying the thermal limit for a particular proportion of the population relies on a dose–response function, we recommend testing at least five different temperatures to infer LT_{80} and TFL_{80}. To make an informed choice of test temperatures, we first suggest using previously estimated CT_{max} (from sources like Bennett et al., 2018; Leiva et al., 2019; Pottier et al., 2022; Weaving et al., 2022) or if resources allow, CT_{max} measurements can be performed (e.g., Hoffmann et al., 2002 for constant temperatures; see; Terblanche et al., 2007 for a ramping protocol). When CTL data is not available, we suggest using existing knowledge of species' geographical ranges (e.g., using the WorldClim2 database, Fick & Hijmans, 2017), which can be adjusted

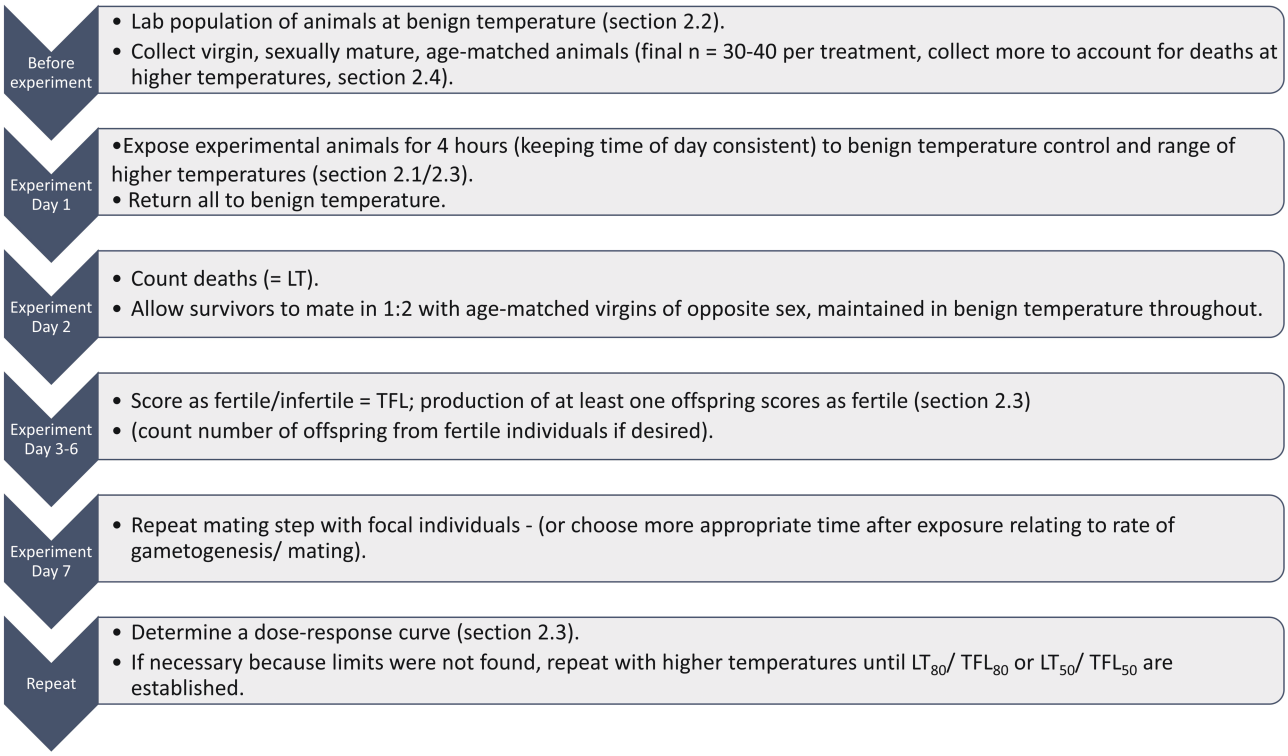

Figure 2. Flowchart of steps to determine TFLs and LTs simultaneously (*Protocol for determining thermal fertility limits* section). Both sexes should be assayed individually, though not necessarily at the same time, given the resources required.

to the ecology of the study species, e.g., using NicheMapR (Kearney & Porter, 2020). Parratt et al. (2021) used the mean maximum air temperature between the years 1970 and 2000 integrated with recorded species ranges to determine the maximum air temperatures species experienced in nature. Subsequent test temperatures would be determined by decreasing in steps of 1–2 °C from the anticipated lethal limit. However, temperature increments should be dictated by the accuracy of the equipment used. To avoid block effects, each experiment should include multiple groups of individuals, with each group exposed to a different temperature. The number of temperatures that can be simultaneously tested is dependent on equipment and organism availability. Multiple rounds may be necessary to find the limits, and each round must include the benign control temperature.

Because our focus is on short extreme heat stress, we recommend a temperature exposure time of 4 hr. During the hottest part of the year, there exists substantial diurnal variation. Typically, according to the diurnal temperature cycle of land surface temperatures, maximum daily temperatures are experienced in the afternoon and last 2–4 hr (Holmes et al., 2015). Our choice of 4 hr therefore simulates this regime and has the added benefit of likely avoiding confounds with humidity due to the short exposure time (Terblanche et al., 2011). It is also practically easy to carry out in a day.

After the 4-hr heat exposure, individuals are returned to their maintenance temperature for 24 hr, regardless of the assay temperature. After 24 hr, survival is scored to determine LT_{80} (Figure 1). Focal survivors are then given mating opportunities with reproductively mature virgins of the opposite sex that have been maintained at the benign temperature (Figure 1). We recommend supplying control non-focal mates at a greater ratio (e.g., 1:2) to avoid the possibility of unintended non-focal individual sterility, unwillingness to mate with heat-stressed focal individuals or other reasons that would confound results. Where the availability of either sex is limiting, then an increase in replication of experimental animals should be prioritized over giving multiple mating partners.

As the gold standard, we recommend measuring the fertility of individuals twice, initially from a mating 24 hr after heat exposure and then at a later time point, to capture delayed sterilization or potential recovery. Previous work on males of 43 *Drosophila* species found more species were sterile at a later time (Parratt et al., 2021). In that case, fertility data was captured after mating over the first 6 days, and the focal males were then provided new virgin females at 7 days after heat exposure treatments and allowed to mate for 24 hr. These timings should be based on traits of the study species, such as remating rate, gametogenesis time and longevity. Alternatively, rather than delayed sterility, fertility may recover at a later time point if physiological repair mechanisms come into play (David et al., 2005; Sales et al., 2021). If resources (time, animals) are limited, then measurement at a one-time point relevant for the given species following heat stress is sufficient. For even richer data, if your organism is amenable, counting the number of offspring can provide finer-scale information about heat-induced effects on fecundity.

To statistically determine the limits, we recommend using a dose–response function as done in previous studies (Parratt et al., 2021; van Heerwaarden & Sgrò, 2021). For instance, one may use the *drm* function in the *drc* package (Ritz et al., 2015) to fit dose–response models using R statistical software or use equivalent models. Our gold-standard protocol uses LT_{80} and TFL_{80} determined for each sex separately, as we assume such a strong reduction would have a severe impact on population

viability (but see *TFL50/LT50* and *Determining the heat stress temperature* sections for alternatives).

Sample size

For the gold-standard approach, we recommend that at least 20, but ideally 30–40, individuals are measured individually to assess TFL at each test temperature up to LT_{80}. For a final sample size of 40, assuming 5 assessment temperatures, the protocol would, therefore, require 200 age-matched individuals of the focal sex. However, the starting number would need to reflect that some individuals will die before mating. Indeed, at LT_{80} only 20% will remain alive to mate, so the starting number for this treatment would need to be 200 in order for the fertility of the surviving 40 to be assessed. Because we recommend assessing sterility at two different time points and allowing mating interaction with multiple control individuals (see *Simultaneous measurement of lethal limits and thermal fertility limits* section), up to 800 individuals of the control sex are necessary to achieve this. We acknowledge that for many species, these sample sizes are unattainable and so below identify ways for reducing the number of individuals needed.

TFL_{50}/LT_{50}

Practical considerations (e.g., resources) may not allow TFL_{80}/LT_{80} to be assessed. It is still valuable to determine how heat stress impacts fertility in such species, and thus, we recommend measuring TFL_{50}/LT_{50}. Measuring the temperature at which 50% of the population dies/is sterile, rather than determining TFL_{80}/LT_{80}, reduces the number of temperature treatments and, therefore, the number of animals assessed and the resources required. Moreover, at LT_{50} only 50% of animals will die before fertility can be measured; hence, for a final sample size of 40 the starting sample size would be 80 (rather than the 200 required at LT_{80} recommended in *Sample size* section). Data so far indicate both TFL_{50} and TFL_{80} predict current thermal ranges and extinction risk (Parratt et al., 2021; van Heerwaarden & Sgrò, 2021).

Protocol for thermal sensitivity of fertility

General considerations

The gold standard (and its alternatives) for measuring TFLs may be impractical or difficult for many organisms, including small terrestrial invertebrates. For example, if sample size is an issue, and/or there are ethical constraints on determining lethal limits, then an alternative is measuring TSF, a measure of proportional loss of fertility (e.g., Baur et al., 2022). Here, organisms are exposed to two temperature treatments—a benign and a heat stress—and the effect on the number of offspring produced (or other proxies of fertility) is quantified (Baur et al., 2022). Reductions in offspring production at a higher temperature indicate how sensitive fertility may be in response to increasing temperatures. Because our goal here is to pragmatically standardize measurements across animal taxa, we recommend TSF be calculated by assessing the effects of a high-temperature heat stress on offspring number after a 4-hr exposure, as with TFLs (Figures 1 and 3). Proxy traits other than offspring number are discussed in *TSF: proxies to offspring counts* section. As with TFL, we recommend mating focal individuals with reproductively mature virgins of the opposite sex that have been maintained at the benign temperature. Because offspring number is being measured, which can be labour-intensive, we recommend supplying only one control non-focal mate to the focal individual, watching whether mating occurs, and then removing the male partner to allow the mated female to oviposit. The length of time for observing mating (or keeping the pair together if mating cannot be observed) or allowing oviposition is species-dependent. The same recommendations about sample size apply here as outlined in the gold-standard protocol (*Sample size* section). Regarding analysis, standard statistical tests to compare two groups, such as a *t*-test or non-parametric Mann–Whitney U test, can be applied. If, in order to achieve the recommended sample size, the experiment has been run with replicate blocks of individuals, then a standard Generalized Linear Model (with appropriate error structure, for example if a Poisson distribution better fits the data) including both temperature and block identity as factors can be used (see Harrison et al., 2018 for a discussion of mixed effect modelling).

Determining the heat stress temperature

To determine TSF, animals are exposed to the benign maintenance temperature and usually one other stressful high temperature (though if resources allow, more than one higher temperature could be used). As the aim of TSF is to assess the change in offspring output between two temperatures and, therefore, be more flexible in application that TFL, we cannot be entirely prescriptive about the choice of temperatures. However, we suggest choosing a stressful temperature estimated from maximum temperatures recorded in the known species geographical range, as discussed in *Simultaneous measurement of lethal limits and thermal fertility limits* section. If these data are not available, but CT_{max} has been measured (e.g., for more than 2,000 species in GlobTherm database, Bennett et al., 2018), then a safe margin below this temperature (e.g., midway between the benign temperature and CT_{max}) could be used. If CT_{max} is not published, and if resources and ethics allow, it can be determined as in *Simultaneous measurement of lethal limits and thermal fertility limits* section.

TSF: proxies to offspring counts

It is sometimes not possible to mate individuals to score offspring production directly. Where collecting sperm or eggs is possible, quantifying impacts on the gametes could be used as a proxy to infer the effect of heat damage on fertility. A recent review summarized evidence that sperm quality and performance measures are indicative of thermal effects on fertility (Wang & Gunderson, 2022).

For males, sperm could be stripped (as in fish, see Supplementary Material 2.1), dissected from the testes or other sperm storage organs, or collected from female reproductive tracts/sperm storage organs after mating. For females, ovaries can be dissected or eggs stripped. Standard assays for live/dead sperm numbers (Eckel et al., 2017; Holman, 2009), sperm traits (van der Horst, 2021), ovariole integrity or DNA damage can then be used (for a general protocol for apoptosis see Sarkissian et al., 2014; for using the TUNEL assay to detect sperm DNA fragmentation see Sharma et al., 2013). Previous studies have measured testes size, sperm number and length (e.g., Sales et al., 2021; Vasudeva et al., 2014), live/dead sperm counts (e.g., for a protocol Eckel et al., 2017; for example Sales et al., 2018), sperm morphology and motility (Hurley et al., 2018; Porcelli et al., 2017), spermatid individualization (Ben-David et al., 2015), and elongation failure

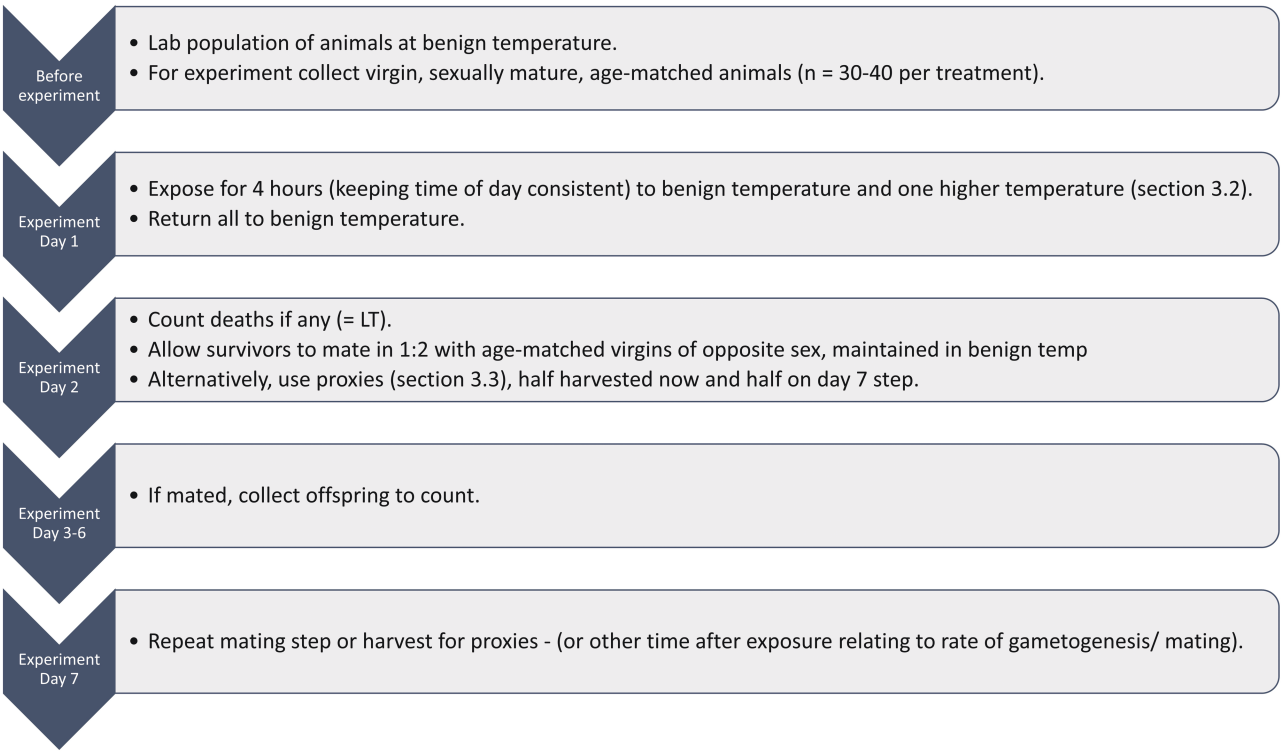

Figure 3. Flowchart of steps to determine TSF (*Protocol for thermal sensitivity of fertility* section). Both sexes should be assayed individually, which could be achieved as one experiment if resources allow.

(Rohmer et al., 2004). Relating to oogenesis, germline stem cell numbers, follicle growth, germline cyst death, and degeneration of vitellogenic follicles can be sensitive to temperature (Gandara & Drummond-Barbosa, 2022). In terms of DNA damage, an increase in double-strand breaks has been observed (Kurhanewicz et al., 2020), and DNA damage in apoptotic cells is measured through the common TUNEL assay (Peña et al., 2019). It is worth considering, however, whether a 4-hr heat shock would be expected to have evident effects on certain traits. For example, morphological traits such as testes size may not change in response to a 4-hr heat shock. While reproductive behavioural traits, such as courtship effort, copulation number or length, may be affected by heat stress, we do not include them as proxies for heat-induced sterility here as our focus is on fertility per se.

A natural history of thermal sensitivity of fertility

General considerations

While controlled laboratory conditions are the gold standard for determining thermal limits to survival and fertility, some organisms are difficult to rear in the laboratory, preventing such estimates. Moreover, it would be relevant to know whether estimates made in the laboratory reflect effects in wild populations (Tratter Kinzner et al., 2019). While two separate laboratory studies on *Drosophila* species demonstrated that TFLs are useful predictors of current species distributions (Parratt et al., 2021; van Heerwaarden & Sgrò, 2021), suggesting there is relevance of laboratory-based measurements for wild populations, field measures are important. Field studies can further our understanding of the effects of extreme heat by measuring responses under real-world thermal variation and associated abiotic variables. In addition, organisms can thermoregulate behaviourally (Muñoz, 2022), which can highlight ways in which animals may avoid thermal extremes. Field studies assessing heat effects on fertility have typically been performed on managed animal populations and domesticated livestock (e.g., egg production; Schou et al., 2021; van Wettere et al., 2021). Thus, there is evidence that naturally occurring heat stress has demonstrable negative effects on fertility.

Here, we provide a protocol for serendipitously assaying the effects of real-world extreme heat on fertility in terrestrial invertebrates (steps outlined in Figure 4). A minimum of three samples will need to be taken within a breeding season, before, during and after the heat event, so that the effect of high temperatures can be compared to a population baseline fertility/ fecundity. We assume that weather forecasts anticipate a period of high temperatures sufficiently far ahead to enable the collection of individuals from a local population before the extreme heat, which can then be compared to individuals collected during and after the extreme heat event. We are avoiding the term "heatwave" as it is difficult to find a consensus on the definition in terms of duration and how far above the long-term mean seasonal temperature a heat event has to be in order to be considered a heatwave. However, we point to the record-breaking high temperatures experienced around the world in the past decade, particularly in the summer of 2023, which, in some areas, lasted for a few days and, in others, multiple weeks. We encourage researchers to "take advantage" of such situations if they occur within the reproductive season of their study species to link these extreme heat events with real-time fertility consequences. If the heat event lasts for an extended period, then we suggest sampling at different points during this time, dependent on

the reproductive schedule of the organism. The short-term and long-term effects of sampling on subsequent effective population size should be considered, especially as we would predict reproduction to be negatively impacted by the high temperatures.

We also provide a Supplementary Material that outlines an intermediate mesocosm/field manipulation approach (Supplementary 1.3). It has been noted that there can be high levels of plasticity in thermal tolerance measured in wild populations (Noer et al., 2022). Our methodology cannot distinguish plasticity from tolerance per se but will give information on the sensitivity to heat stress in natural populations by comparing temporally spaced samples. However, as both lab and field data accumulate, if field data show limited thermal effects on fertility, whereas lab data show strong effects, then it may be possible to infer some aspects of plasticity are likely involved in natural populations.

Sampling from the wild

Our recommended methodology for this approach is to capture wild animals and bring them into the laboratory to assess their reproductive output. When collecting from the wild and assaying in the laboratory, it is generally recommended to acclimatize recently caught animals to benign temperatures prior to measuring any physiological or behavioural trait (e.g., Moretti et al., 2017). For our purposes, we recommend that animals be maintained in the laboratory at a standard benign temperature and that fertility is measured immediately, without acclimation. This is because fertility recovery at permissive temperatures could occur, obscuring any immediate effects of heat stress on fertility. When using wild-caught individuals, various confounding factors may influence the results. These include age, mating status and previous heat exposure (e.g., hardening), thus a record of environmental data (at a minimum, the mean and maximum temperature the preceding week, but ideally using data loggers for fine-scale data capture) before and on the day of sampling should be provided. Depending on the species, it may be possible to estimate age and, more likely for females, mating status, both of which may modify responses to heat (Baur et al., 2022; Vasudeva et al., 2021; Walsh et al., 2022). As with the laboratory protocols, care should be taken to choose sample sizes to adjust for this expected variation in wild-caught animals (see below). In addition, if the prior baseline sample is taken too far in advance of the heat event, then fertility differences might reflect seasonality or age-dependent changes rather than the extreme temperature change itself. Such factors should be carefully and openly reported, and caution should be taken when interpreting the results.

Measuring reproductive output from wild-caught individuals in this way might allow estimation of TFL. That is, if 80% of those individuals are found to be infertile, then they have experienced a temperature akin to TFL_{80}. This would require at least 20 individuals to be captured and survive long enough to produce offspring (meaning 16 would be infertile if TFL_{80} had been reached). However, this must be compared to a baseline; for example, if in animals collected after benign temperatures, 20% are found to be infertile, then this must be taken into account. That is, in a sample of 20 individuals, 4 would not produce offspring, so if 16 were infertile, this would be TFL_{75}, not TFL_{80}. To report a TFL in this way, the proportion of infertile individuals can be adjusted to the baseline in a manner similar to assessing competitive paternity share (P1/P2) (see Boorman & Parker, 1976). That is, the proportion of sterile individuals is equal to:

$$1 - [(\text{after sample sterile}/\text{ after sample total}) /(\text{before sample sterile}/\text{before sample total})]$$

The number of fertile/infertile individuals across the three time points could be compared with a contingency chi-squared test (or Fisher's exact test if expected values are less than 5). However, given the sample size requirements above, it is likely that a measure of TSF is more attainable, and we would suggest a minimum sample size of 10 per sex per time point, though 20–30 would be ideal. Whether it is possible to measure sex-specific effects of heat stress on fertility depends on whether control-temperature individuals from a laboratory population are available. Below, we outline sex-specific approaches. Whether offspring number or a fertility proxy is measured, this can be compared across the three time points using GLMs (with appropriate error structures). Ambient temperatures should be reported but cannot be used as a co-variate because of the collinearity with the time point.

Females

The ideal scenario would be to capture virgin females from the wild and mate them with laboratory-reared standardized males. However, it is unlikely that the majority of adult females will be virgins. Previous laboratory-based work in *Drosophila* spp. has shown reduced reproduction of females after heat stress due to heat damage of stored sperm within the females (Walsh et al., 2022). Given that females of many species store sperm and heat damage to sperm is common, delineating sex-specific effects of thermal stress based on offspring numbers produced by wild-caught females will be difficult. However, it will allow for the estimation of population-level consequences of heat stress. If researchers are interested in female-specific fertility effects due to natural extreme heat, then using proxy oogenesis measures as described in *TSF: proxies to offspring counts* section is recommended, and knowledge about a female's reproductive cycle/mating status in relation to the temporal dynamics of the heat event is necessary.

Males

Measuring offspring output from wild-caught males necessitates having a set of standardized virgin females for mating. However, it is at least clear that any alteration of reproduction arising from those matings must be a male effect. Changes in fertility in response to extreme heat can be assessed by comparing offspring production of males collected before, during and after the event. However, if a laboratory culture is not available, then, as with females, proxies to estimate the effects of natural extreme heat on wild-caught male fertility can also be used (see *TSF: proxies to offspring counts* section).

No sex-specific delineation

If possible, determining the sex-specific effect of extreme heat on fertility is recommended as such data will allow for more fine-grained models forecasting population viability to future climate change. However, being pragmatic, simply knowing the extent of effects on the population is important too. If researchers cannot estimate sex-specific effects as outlined in *Females and males* sections, and if wild-caught females will produce offspring either in the laboratory or in a managed (or monitored) environment, then simply collecting females

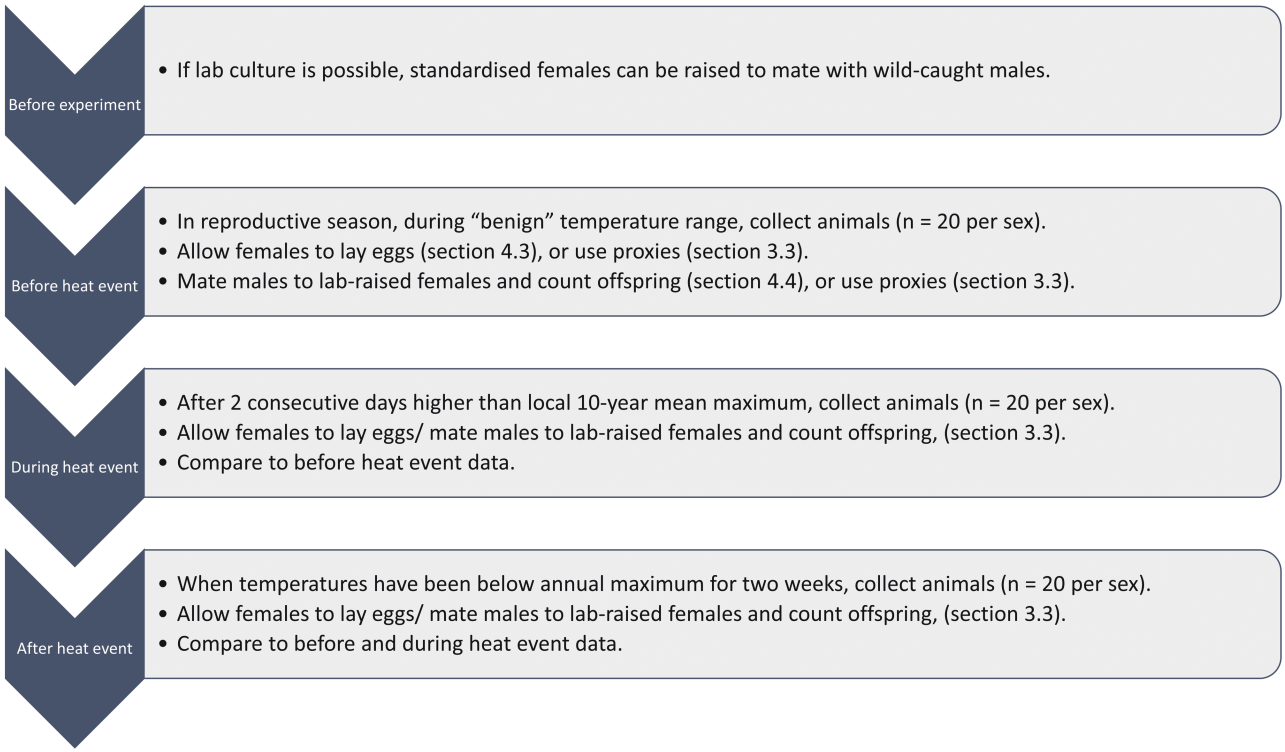

Figure 4. Flowchart of steps to determine a natural history of thermal fertility (*A natural history of thermal sensitivity of fertility* section).

and comparing offspring output before, during and after an extreme heat event would be useful. However, without being able to mate with laboratory-reared males, it is difficult to determine whether offspring reduction is due to heat stress or simply sperm limitation because females have not mated recently enough. To establish a baseline for comparison, the number of offspring from females caught before the extreme heat should be monitored over the timeframe of the extreme temperatures, if possible.

Data recording and accessibility

Researchers must collate and publish their data, ensuring they are readily findable, accessible, interoperable and reusable (see FAIR Data Principles https://www.go-fair.org/fair-principles/), including the experimental design metadata (Table 1). There is an accelerating move towards compiling open-access databases that enable the community to identify broad ecological and evolutionary patterns. For example, the GlobTherm database (Bennett et al., 2018) collated experimental data on thermal tolerances (CT_{min}, CT_{max}) across a broad range of taxa, subsequently identifying traits associated with thermal tolerance, which deepened understanding of future consequences of increasing temperature to biodiversity (Bennett et al., 2021). Databases are also useful for meta-analysis, but measurement error (and other associated data, i.e., sample size; see Table 1) must be included for the trait of interest. Some traits, like TFL, are point estimates without error, but others, like TSF, will include measurement error. We aim to develop a database for thermal fertility data in the near future, based largely on our standardized protocol. Such a resource would be a valuable basis for (a) future comparative investigations on thermal fertility effects, (b) predictions of future species vulnerability and distribution changes in ecophysiological models, (c) assessments of the relative impact of different physiological limits on these forecasts, and (d) estimates of phylogenetic inertia and constraints. With a view to such a future database, we provide in Table 1 a checklist of minimum information that would be included. We suggest that this checklist could help to clarify experimental design as well as the data that need to be recorded when following the protocol, and so we encourage researchers to consider it before starting their data collection.

Discussion

In proposing these protocols and databases, we hope to encourage the rapid collection of standardized data to fill critical knowledge gaps needed to identify the vulnerability of fertility to climate change. Our goal was to be pragmatic to encourage the wide adoption of techniques and to encourage rapid data collection. The TFL approach is based on the successful CT_{max} paradigm, which has been widely adopted for forecasting climate change responses. Indeed, the application of TFLs has already shown its link to current species distributions and likely consequences for future range limits (Parratt et al., 2021; van Heerwaarden & Sgrò, 2021).

To provide a standardized method for determining TFLs and TSFs, useful for subsequent application to improving species forecasts of response, which will be to climate change and to perform cross-species comparisons (i.e., as is done for CT_{max} scores) and to mimic increasing frequency and intensity of heat stress, we advocate for a 4-hr heat stress applied in the hottest part of the day (see also Parratt et al., 2021). These data will provide a baseline for thermal effects on fertility. Subsequent experiments can add more ecological and evolutionary realism, such as exposure to longer durations or multiple-day exposure during development (see

Supplementary 1.4), across fluctuating temperatures (van Heerwaarden & Sgrò, 2021), whether previous exposure mitigates (hardens) or exacerbates effects, etc. Such subsequent studies will improve forecasting how increasing mean and variability of global temperatures may impact population viability.

We also argue that the gold-standard method is to determine TFLs because it is more aligned with the wealth of research using CT_{max} to predict ecological and evolutionary consequences of increasing temperatures, and thus, comparisons between modelling output using CT_{max} vs. TFLs can be performed. However, we also offer protocols based on TSF assessments (e.g., Baur et al., 2022), which may provide more fine-grained estimates of the effect of heat stress on population size. The extent to which these different measures of thermal effects on fertility (TFL as an absolute fertility measure vs. TSF as a quantitative measure) impact subsequent population persistence will need to be modelled. Where possible, researchers may want to do both, i.e., calculate both fertility effects and progeny number effects in the same population, which can be done by counting the offspring of the remaining fertile individuals in the TFL protocol (Rodrigues et al., 2022). Keep in mind however that offspring counts increase the time and resources necessary to produce data.

In cases where either estimations of TFL or TSF through offspring counts are not possible, we suggest proxies based on gamete performance. Such data can also help pinpoint the mechanisms by which heat stress affects fertility and those parts of the reproductive process that are more sensitive to heat damage. In turn, this mechanistic understanding may help to predict the extent to which heat-induced sterility is either temporary or permanent. Further key questions are whether individuals can recover fertility after sterilizing temperatures, whether acclimation (i.e., prior experience of heat) can increase TFLs, and whether TFLs are evolvable. We have not outlined recovery experiments (e.g., Canal Domenech & Fricke, 2022; Sales et al., 2021). However, determining whether full or partial recovery occurs will also help to forecast the overall cost of increasing temperatures to population viability. Other factors that may either reduce or exacerbate heat stress effects on fertility are, for instance, age/life stage, previous mating status, or body size etc., all of which can be assayed in subsequent experiments manipulating these factors, adding more ecological realism to models forecasting climate change consequences to species.

Overall, our protocols address the outstanding questions we outlined in the Introduction section, and we discuss additions to this work that expand understanding of the role of increasing temperatures on fertility and potential subsequent consequences for species distributions and population persistence. We invite researchers to apply these standardized protocols broadly across animal taxa, including those we have not been able to specifically include due to our range of expertise. While we acknowledge such standard protocols are the starting point of ecological evaluations and welcome others to extend these ideas, the urgency of collecting fertility data cannot be overstated, as it is crucial to understanding the biological responses of different species under global warming. Ultimately, expanding the taxa studied and collectively enhancing our understanding of the vital aspect fertility plays in species' thermal biology will profoundly reshape our understanding of organisms' response to increasing temperatures.

Supplementary material

Supplementary material is available at *Journal of Evolutionary Biology* online.

Author contributions

Amanda Bretman (Conceptualization [lead], Funding acquisition [lead], Methodology [lead], Visualization [equal], Writing—original draft [lead], Writing—review & editing [equal]), Claudia Fricke (Conceptualization [lead], Funding acquisition [lead], Methodology [lead], Writing—original draft [lead], Writing—review & editing [equal]), Julian Baur (Methodology [supporting], Writing—original draft [supporting], Writing—review & editing [equal]), David Berger (Writing—review & editing [equal]), Merel Breedveld (Methodology [supporting], Writing—original draft [supporting], Writing—review & editing [equal]), Berta Canal (Methodology [supporting], Writing—original draft [supporting], Writing—review & editing [equal]), Diego Dierick (Methodology [supporting], Writing—original draft [supporting], Writing—review & editing [equal]), Szymon Drobniak (Methodology [supporting], Writing—original draft [supporting], Writing—review & editing [equal]), Jacintha Ellers (Methodology [supporting], Writing—original draft [supporting], Writing—review & editing [equal]), Sinead English (Writing—review & editing [equal]), Clelia Gasparini (Methodology [supporting], Writing—original draft [supporting], Writing—review & editing [equal]), Graziella Iossa (Methodology [supporting], Writing—original draft [supporting], Writing—review & editing [equal]), Malgorzata Lagisz (Visualization [lead], Writing—review & editing [equal]), Shinichi Nakagawa (Writing—review & editing [equal]), Daniel Noble (Methodology [supporting], Writing—original draft [supporting], Writing—review & editing [equal]), Patrice Pottier (Writing—review & editing [equal]), Steven Ramm (Methodology [supporting], Writing—original draft [supporting], Writing—review & editing [equal]), Melissah Rowe (Methodology [supporting], Writing—original draft [supporting], Writing—review & editing [equal]), Mads Schou (Writing—review & editing [equal]), Eva Schultner (Methodology [supporting], Writing—original draft [supporting], Writing—review & editing [equal]), Pedro Simões (Methodology [supporting], Writing—original draft [supporting], Writing—review & editing [equal]), Paula Stockley (Methodology [supporting], Writing—original draft [supporting], Writing—review & editing [equal]), Ramakrishnan Vasudeva (Methodology [supporting], Visualization [supporting], Writing—original draft [supporting], Writing—review & editing [equal]), Hester Weaving (Writing—review & editing [equal]), Tom Price (Conceptualization [lead], Funding acquisition [lead], Methodology [lead], Writing—original draft [lead], Writing—review & editing [equal]), and Rhonda R. Snook (Conceptualization [lead], Funding acquisition [lead], Methodology [lead], Writing—original draft [lead], Writing—review & editing [equal])

Funding

This work was funded by the European Society for Evolutionary Biology which funds a Special Topic Network on Evolutionary Ecology of Thermal Fertility Limits (to C.F., A.B., R.R.S. and T.A.R.P.), by the Natural Environment

Research Council (NE/P002692/1 to T.A.R.P., A.B. and R.R.S.), the Biotechnology and Biosciences Research Council (BB/W016753/1 to A.B., T.A.R.P. and R.R.S.), Australian Research Council (DE180100202), National Science Centre of Poland (to S.M.D.) and a Heisenberg fellowship from the Deutsche Forschungsgemeinschaft to C.F. (FR 2973/11-1).

Acknowledgments

A.B., C.F., T.A.R.P., and R.R.S. contributed equally to this work.

Conflicts of interest

None declared.

References

Baur, J., Jagusch, D., Michalak, P., … Berger, D. (2022). The mating system affects the temperature sensitivity of male and female fertility. *Functional Ecology*, *36*(1), 92–106. https://doi.org/10.1111/1365-2435.13952

Ben-David, G., Miller, E., & Steinhauer, J. (2015). *Drosophila* spermatid individualization is sensitive to temperature and fatty acid metabolism. *Spermatogenesis*, *5*(1), e1006089. https://doi.org/10.1080/21565562.2015.1006089

Bennett, J. M., Calosi, P., Clusella-Trullas, S., … Morales-Castilla, I. (2018). GlobTherm, a global database on thermal tolerances for aquatic and terrestrial organisms. *Scientific Data*, *5*, 180022. https://doi.org/10.1038/sdata.2018.22

Bennett, J. M., Sunday, J., Calosi, P., … Olalla-Tárraga, M. A. (2021). The evolution of critical thermal limits of life on Earth. *Nature Communications*, *12*(1), 1198. https://doi.org/10.1038/s41467-021-21263-8

Boorman, E., & Parker, G. A. (1976). Sperm (ejaculate) competition in *Drosophila melanogaster*, and the reproductive value of females to males in relation to female age and mating status. *Ecological Entomology*, *1*(3), 145–155. https://doi.org/10.1111/j.1365-2311.1976.tb01217.x

Bush, A., Mokany, K., Catullo, R., … Ferrier, S. (2016). Incorporating evolutionary adaptation in species distribution modelling reduces projected vulnerability to climate change. *Ecology Letters*, *19*(12), 1468–1478. https://doi.org/10.1111/ele.12696

Calvin, K., Dasgupta, D., Krinner, G., … Péan, C. (2023). In Core Writing Team, H. Lee, & J. Romero (Eds.), *IPCC, 2023: Climate change 2023: Synthesis report. Contribution of working groups I, II and III to the sixth assessment report of the intergovernmental panel on climate change*. IPCC, Intergovernmental Panel on Climate Change (IPCC). https://doi.org/10.59327/IPCC/AR6-9789291691647

Canal Domenech, B., & Fricke, C. (2022). Recovery from heat-induced infertility—A study of reproductive tissue responses and fitness consequences in male *Drosophila melanogaster*. *Ecology and Evolution*, *12*, e9563. https://doi.org/10.1002/ece3.9563

David, J. R., Araripe, L. O., Chakir, M., … Moreteau, B. (2005). Male sterility at extreme temperatures: A significant but neglected phenomenon for understanding *Drosophila* climatic adaptations. *Journal of Evolutionary Biology*, *18*(4), 838–846. https://doi.org/10.1111/j.1420-9101.2005.00914.x

Deutsch, C. A., Tewksbury, J. J., Huey, R. B., … Martin, P. R. (2008). Impacts of climate warming on terrestrial ectotherms across latitude. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(18), 6668–6672. https://doi.org/10.1073/pnas.0709472105

Dougherty, L. R., Frost, F., Maenpaa, M. I., … Price, T. A. R. (2024). A systematic map of studies testing the relationship between temperature and animal reproduction. *Ecological Solutions and Evidence*, *5*, e12303.

Eckel, B. A., Guo, R., & Reinhardt, K. (2017). More pitfalls with sperm viability staining and a viability-based stress test to characterize sperm quality. *Frontiers in Ecology and Evolution*, *5*, 165. https://doi.org/10.3389/fevo.2017.00165

Eisenhauer, N., Bonn, A., & Guerra, C. A. (2019). Recognizing the quiet extinction of invertebrates. *Nature Communications*, *10*, 50. https://doi.org/10.1038/s41467-018-07916-1

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Gandara, A. C. P., & Drummond-Barbosa, D. (2022). Warm and cold temperatures have distinct germline stem cell lineage effects during *Drosophila* oogenesis. *Development*, *149*(5), dev200149. https://doi.org/10.1242/dev.200149

Gouveia, S. F., Hortal, J., Tejedo, M., … Diniz-Filho, J. A. F. (2014). Climatic niche at physiological and macroecological scales: The thermal tolerance–geographical range interface and niche dimensionality. *Global Ecology and Biogeography*, *23*(4), 446–456. https://doi.org/10.1111/geb.12114

Haelewaters, D., Hofmann, T. A., & Romero-Olivares, A. L. (2021). Ten simple rules for Global North researchers to stop perpetuating helicopter research in the Global South. *PLoS Computational Biology*, *17*(8), e1009277. https://doi.org/10.1371/journal.pcbi.1009277

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., … Inger, R. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, *6*, e4794. https://doi.org/10.7717/peerj.4794

Hector, T. E., Gehman, A. -L. M., & King, K. C. (2023). Infection burdens and virulence under heat stress: Ecological and evolutionary considerations. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *378*, 20220018. https://doi.org/10.1098/rstb.2022.0018

Hoffmann, A. A., Anderson, A., & Hallas, R. (2002). Opposing clines for high and low temperature resistance in *Drosophila melanogaster*. *Ecology Letters*, *5*(5), 614–618. https://doi.org/10.1046/j.1461-0248.2002.00367.x

Hoffmann, A. A., & Ross, P. A. (2018). Rates and patterns of laboratory adaptation in (mostly) insects. *Journal of Economic Entomology*, *111*(2), 501–509. https://doi.org/10.1093/jee/toy024

Holman, L. (2009). Sperm viability staining in ecology and evolution: Potential pitfalls. *Behavior, Ecology and Sociobiology*, *63*(11), 1679–1688. https://doi.org/10.1007/s00265-009-0816-4

Holmes, T. R. H., Crow, W. T., Hain, C., … Kustas, W. P. (2015). Diurnal temperature cycle as observed by thermal infrared and microwave radiometers. *Remote Sensing of Environment*, *158*, 110–125. https://doi.org/10.1016/j.rse.2014.10.031

Hurley, L. L., McDiarmid, C. S., Friesen, C. R., … Rowe, M. (2018). Experimental heatwaves negatively impact sperm quality in the zebra finch. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *285*(1871), 20172547. https://doi.org/10.1098/rspb.2017.2547

Iossa, G. (2019). Sex-specific differences in thermal fertility limits. *Trends in Ecology and Evolution*, *34*(6), 490–492. https://doi.org/10.1016/j.tree.2019.02.016

Kearney, M. R., & Porter, W. P. (2020). NicheMapR—An R package for biophysical modelling: the ectotherm and Dynamic Energy Budget models. *Ecography*, *43*(1), 85–96. https://doi.org/10.1111/ecog.04680

Kellermann, V., Overgaard, J., Hoffmann, A. A., … Loeschcke, V. (2012). Upper thermal limits of *Drosophila* are linked to species distributions and strongly constrained phylogenetically. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(40), 16228–16233. https://doi.org/10.1073/pnas.1207553109

Kurhanewicz, N. A., Dinwiddie, D., Bush, Z. D., & Libuda, D. E. (2020). Elevated temperatures cause transposon-associated DNA damage in *C. elegans* spermatocytes. *Current Biology*, *30*(24), 5007–5017. e4. https://doi.org/10.1016/j.cub.2020.09.050

Leiva, F. P., Calosi, P., & Verberk, W. C. E. P. (2019). Scaling of thermal tolerance with body mass and genome size in ectotherms: A comparison between water- and air-breathers. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *374*(1778), 20190035. https://doi.org/10.1098/rstb.2019.0035

Leong, C. -M., Tsang, T. P. N., & Guénard, B. (2022). Testing the reliability and ecological implications of ramping rates in the measurement of Critical Thermal maximum. *PLoS One*, *17*(3), e0265361. https://doi.org/10.1371/journal.pone.0265361

Lutterschmidt, W. I., & Hutchison, V. H. (1997). The critical thermal maximum: History and critique. *Canadian Journal of Zoology*, *75*(10), 1561–1574. https://doi.org/10.1139/z97-783

Moretti, M., Dias, A. T. C., Bello, F., … Berg, M. P. (2017). Handbook of protocols for standardized measurement of terrestrial invertebrate functional traits. *Functional Ecology*, *31*, 558–567. https://doi.org/10.1111/1365-2435.12776

Morgan, R., Sundin, J., Finnøen, M. H., … Jutfelt, F. (2019). Are model organisms representative for climate change research? Testing thermal tolerance in wild and laboratory zebrafish populations. *Conservation Physiology*, *7*(1), coz036. https://doi.org/10.1093/conphys/coz036

Muñoz, M. M. (2022). The Bogert effect, a factor in evolution. *Evolution*, *76*(S1), 49–66. https://doi.org/10.1111/evo.14388

Murali, G., Iwamura, T., Meiri, S., & Roll, U. (2023). Future temperature extremes threaten land vertebrates. *Nature*, *615*(7952), 461–467. https://doi.org/10.1038/s41586-022-05606-z

Noer, N. K., Ørsted, M., Schiffer, M., … Kristensen, T. N. (2022). Into the wild—A field study on the evolutionary and ecological importance of thermal plasticity in ectotherms across temperate and tropical regions. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *377*(1846), 20210004. https://doi.org/10.1098/rstb.2021.0004

Parratt, S. R., Walsh, B. S., Metelmann, S., … Price, T. A. R. (2021). Temperatures that sterilize males better match global species distributions than lethal temperatures. *Nature Climate Change*, *11*(6), 481–484. https://doi.org/10.1038/s41558-021-01047-0

Pecl, G. T., Araújo, M. B., Bell, J. D., … Williams, S. E. (2017). Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science*, *355*(6332), eaai9214. https://doi.org/10.1126/science.aai9214

Peña, S. T., Stone, F., Gummow, B., … Paris, D. B. B. P. (2019). Tropical summer induces DNA fragmentation in boar spermatozoa: Implications for evaluating seasonal infertility. *Reproduction Fertility and Development*, *31*, 590–601. https://doi.org/10.1071/RD18159

Pinsky, M. L., Eikeset, A. M., McCauley, D. J., … Sunday, J. M. (2019). Greater vulnerability to warming of marine versus terrestrial ectotherms. *Nature*, *569*(7754), 108–111. https://doi.org/10.1038/s41586-019-1132-4

Porcelli, D., Gaston, K. J., Butlin, R. K., & Snook, R. R. (2017). Local adaptation of reproductive performance during thermal stress. *Journal of Evolutionary Biology*, *30*(2), 422–429. https://doi.org/10.1111/jeb.13018

Porras, M. F., Navas, C. A., Marden, J. H., … Carlo, T. A. (2020). Enhanced heat tolerance of viral-infected aphids leads to niche expansion and reduced interspecific competition. *Nature Communications*, *11*(1), 1184. https://doi.org/10.1038/s41467-020-14953-2

Pottier, P., Lin, H. -Y., Oh, R. R. Y., … Nakagawa, S. (2022). A comprehensive database of amphibian heat tolerance. *Scientific Data*, *9*(1), 600. https://doi.org/10.1038/s41597-022-01704-9

Proctor, H. (1998). Indirect sperm transfer in arthropods: Behavioral and evolutionary trends. *Annual Review of Entomology*, *43*, 153–174. https://doi.org/10.1146/annurev.ento.43.1.153

Riddell, E. A., Mutanen, M., & Ghalambor, C. K. (2023). Hydric effects on thermal tolerances influence climate vulnerability in a high-latitude beetle. *Global Change Biology*, *29*(18), 5184–5198. https://doi.org/10.1111/gcb.16830

Ritz, C., Baty, F., Streibig, J. C., & Gerhard, D. (2015). Dose-response analysis using R. *PLoS One*, *10*(12), e0146021. https://doi.org/10.1371/journal.pone.0146021

Rodrigues, L. R., Zwoinska, M. K., Wiberg, R. A. W., & Snook, R. R. (2022). The genetic basis and adult reproductive consequences of developmental thermal plasticity. *Journal of Animal Ecology*, *91*(6), 1119–1134. https://doi.org/10.1111/1365-2656.13664

Rohmer, C., David, J. R., Moreteau, B., & Joly, D. (2004). Heat induced male sterility in *Drosophila melanogaster*: Adaptive genetic variations among geographic populations and role of the Y chromosome. *Journal of Experimental Biology*, *207*(Pt 16), 2735–2743. https://doi.org/10.1242/jeb.01087

Sales, K., Vasudeva, R., Dickinson, M. E., … Gage, M. J. G. (2018). Experimental heatwaves compromise sperm function and cause transgenerational damage in a model insect. *Nature Communications*, *9*(1), 4771. https://doi.org/10.1038/s41467-018-07273-z

Sales, K., Vasudeva, R., & Gage, M. J. G. (2021). Fertility and mortality impacts of thermal stress from experimental heatwaves on different life stages and their recovery in a model insect. *Royal Society Open Science*, *8*(3), 201717. https://doi.org/10.1098/rsos.201717

Sarkissian, T., Timmons, A., Arya, R., … White, K. (2014). Detecting apoptosis in *Drosophila* tissues and cells. *Methods*, *68*(1), 89–96. https://doi.org/10.1016/j.ymeth.2014.02.033

Schou, M. F., Bonato, M., Engelbrecht, A., … Cornwallis, C. K. (2021). Extreme temperatures compromise male and female fertility in a large desert bird. *Nature Communications*, *12*(1), 666. https://doi.org/10.1038/s41467-021-20937-7

Sharma, R., Masaki, J., & Agarwal, A. (2013). Sperm DNA fragmentation analysis using the TUNEL assay. In D. T. Carrell & K. I. Aston (Eds.), *Spermatogenesis, methods in molecular biology* (pp. 121–136). Humana Press. https://doi.org/10.1007/978-1-62703-038-0_12

Stillman, J. H. (2019). Heat waves, the new normal: Summertime temperature extremes will impact animals, ecosystems, and human communities. *Physiology (Bethesda, Md.)*, *34*(2), 86–100. https://doi.org/10.1152/physiol.00040.2018

Sunday, J. M., Bates, A. E., & Dulvy, N. K. (2012). Thermal tolerance and the global redistribution of animals. *Nature Climate Change*, *2*(9), 686–690. https://doi.org/10.1038/nclimate1539

Terblanche, J. S., Deere, J. A., Clusella-Trullas, S., … Chown, S. L. (2007). Critical thermal limits depend on methodological context. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *274*(1628), 2935–2942. https://doi.org/10.1098/rspb.2007.0985

Terblanche, J. S., Hoffmann, A. A., Mitchell, K. A., … Chown, S. L. (2011). Ecologically relevant measures of tolerance to potentially lethal temperatures. *Journal of Experimental Biology*, *214*(Pt 22), 3713–3725. https://doi.org/10.1242/jeb.061283

Tratter Kinzner, M., Kinzner, M. -C., Kaufmann, R., … Steiner, F. M. (2019). Is temperature preference in the laboratory ecologically relevant for the field? The case of *Drosophila nigrosparsa*. *Global Ecology and Conservation*, *18*, e00638. https://doi.org/10.1016/j.gecco.2019.e00638

Urban, M. C., Bocedi, G., Hendry, A. P., … Travis, J. M. J. (2016). Improving the forecast for biodiversity under climate change. *Science*, *353*(6304), aad8466. https://doi.org/10.1126/science.aad8466

van der Horst, G. (2021). Processes and data management of computer-aided sperm analysis in human and animal spermatozoa. In L. Björndahl, J. Flanagan, R. Holmberg, & U. Kvist (Eds.), *XIIIth international symposium on spermatology* (pp. 199–210). Springer International Publishing. https://doi.org/10.1007/978-3-030-66292-9_27

van Heerwaarden, B., & Sgrò, C. M. (2021). Male fertility thermal limits predict vulnerability to climate warming. *Nature Communications*, *12*(1), 2214. https://doi.org/10.1038/s41467-021-22546-w

van Wettere, W. H. E. J., Kind, K. L., Gatford, K. L., … Walker, S. K. (2021). Review of the impact of heat stress on reproductive

performance of sheep. *Journal of Animal Science and Biotechnology*, *12*(1), 26. https://doi.org/10.1186/s40104-020-00537-z

Vasudeva, R., Deeming, D. C., & Eady, P. E. (2014). Developmental temperature affects the expression of ejaculatory traits and the outcome of sperm competition in *Callosobruchus maculatus*. *Journal of Evolutionary Biology*, *27*(9), 1811–1818. https://doi.org/10.1111/jeb.12431

Vasudeva, R., Deeming, D. C., & Eady, P. E. (2021). Age-specific sensitivity of sperm length and testes size to developmental temperature in the bruchid beetle. *Journal of Zoology*, *314*(4), 285–295. https://doi.org/10.1111/jzo.12884

Walsh, B. S., Parratt, S. R., Atkinson, D., … Price, T. A. R. (2019). Integrated approaches to studying male and female thermal fertility limits. *Trends in Ecology and Evolution*, *34*(6), 492–493. https://doi.org/10.1016/j.tree.2019.03.005

Walsh, B. S., Parratt, S. R., Hoffmann, A. A., … Price, T. A. R. (2019). The impact of climate change on fertility. *Trends in Ecology and Evolution*, *34*(3), 249–259. https://doi.org/10.1016/j.tree.2018.12.002

Walsh, B. S., Parratt, S. R., Snook, R. R., … Price, T. A. R. (2022). Female fruit flies cannot protect stored sperm from high temperature damage. *Journal of Thermal Biology*, *105*, 103209. https://doi.org/10.1016/j.jtherbio.2022.103209

Wang, W. W. -Y., & Gunderson, A. R. (2022). The physiological and evolutionary ecology of sperm thermal performance. *Frontiers in Physiology*, *13*, 754830. https://doi.org/10.3389/fphys.2022.754830

Weaving, H., Terblanche, J. S., Pottier, P., & English, S. (2022). Meta-analysis reveals weak but pervasive plasticity in insect thermal limits. *Nature Communications*, *13*(1), 5292. https://doi.org/10.1038/s41467-022-32953-2

Witte, H., & Döring, D. (1999). Canalized pathways of change and constraints in the evolution of reproductive modes of microarthropods. In: J. Bruin, L. P. S. van der Geest, & M. W. Sabelis (Eds.), *Ecology and evolution of the Acari*. Series Entomologica (Vol. *55*). Springer. https://doi.org/10.1007/978-94-017-1343-6_2