2024

Computer vision enabled approaches to predicting responses to fluctuating temperatures in aquatic embryos

Ibbini, Ziad

UNIVERSITY OF
PLYMOUTH

COMPUTER VISION ENABLED APPROACHES TO PREDICTING RESPONSES TO FLUCTUATING TEMPERATURES IN AQUATIC EMBRYOS

by

Ziad Fayez Ibbini

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Biological and Marine Sciences

August 2024

Acknowledgements

There are many people whom I need to thank for all their help during my PhD.

First, thank you to my Director of Studies, Dr Oliver Tills, who has provided immense support and advice throughout my PhD, and has given me the environment and opportunities to realise my passion for computational biology. Also, thank you to Dr Manuela Truebano Garcia, my second supervisor, who has kept me directed and focused throughout my PhD. I would also like to thank my third supervisor, Dr John Bishop, for his expertise on working with ascidians, and for his input into my manuscripts which has been greatly valued.

On a more formal note, thank you to the School of Biological and Marine Sciences for funding my PhD studentship, the University of Plymouth and Marine Biological Association for use of the laboratory facilities, and Plymouth Science Park for supporting me with space and facilities.

I would also like to thank other staff and collaborators, including Prof. John Spicer, Dr Jamie McCoy, Prof. Simon Rundle and Prof. Enrico Rezende, who have provided helpful insight and advice at various stages of my PhD. Thank you to all the technical staff, including Marie Palmer, Chris Woods, Dr Charlotte Crowther and Andrew Atfield, for all their help and advice in setting up and maintaining laboratory cultures. Many thanks also to the other PhD students and other staff within MBERC.

Lastly, I would like to thank my wife, Daria Ibbini, for her tireless support and patience over the years, and for keeping me sane through the toughest times of my PhD, and thank you to my family, for their continued support and words of encouragement throughout. You have all had a part in shaping who I am today.

Author's declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Doctoral College Quality Sub-Committee.

Work submitted for this research degree at the University of Plymouth has not formed part of any other degree either at the University of Plymouth or at another establishment.

This study was financed with the aid of a studentship from the School of Biological and Marine Sciences, University of Plymouth.

A programme of advanced study was undertaken, which included the taught module BIO5131 Postgraduate Research Skills.

Publications (or public presentation of creative research outputs):

Tills, O., Spicer, J.I., Ibbini, Z. and Rundle, S.D., 2021. Spectral phenotyping of embryonic development reveals integrative thermodynamic responses. *BMC bioinformatics*, *22*(1), p.232.

Ibbini, Z., Spicer, J.I., Truebano, M., Bishop, J. and Tills, O., 2022. HeartCV: a tool for transferrable, automated measurement of heart rate and heart rate variability in transparent animals. *Journal of Experimental Biology*, *225*(19), p.jeb244729.

McCoy, J.C., Spicer, J.I., Ibbini, Z. and Tills, O., 2023. Phenomics as an approach to Comparative Developmental Physiology. *Frontiers in Physiology*, *14*, p.1229500.

Ibbini, Z., Truebano, M., Spicer, J.I., McCoy, J. and Tills, O., 2024. Dev-ResNet: automated developmental event detection using deep learning. *Journal of Experimental Biology*, *227*(10).

Presentations at conferences:

SEB Annual conference 2021, "HeartCV: Automated, noninvasive quantification of cardiac traits from transparent animals", Poster presentation online, July 2021.

SICB Annual conference 2023, "HeartCV: A tool for transferrable, automated measurement of cardiac function in transparent animals", Poster presentation, January 2023.

SEB Annual conference 2023, "Predicting developmental responses to fluctuating temperatures in *Lymnaea stagnalis*", July 2023.

Word count of main body of thesis: 35,132

Signed: 

Date: 14/08/2024

Abstract

Computer vision enabled approaches to predicting responses to fluctuating temperatures in aquatic embryos

Ziad Fayez Ibbini

Fluctuations in temperature are prevalent within environments inhabited by many ectotherms, and climate change continues to rapidly alter the thermal regimes of these environments. Early developmental stages are often the most thermally sensitive life stage, and thus can drive a species' vulnerability to climate change. Our current understanding of the effects of fluctuating temperatures (FTs) is predominately based upon comparisons of responses to constant temperatures (CTs). Relating CT responses to those of FTs is typically achieved through the use of thermal performance curves, which are integrated over time to estimate cumulative performance and thus a developmental endpoint; an approach often termed rate summation. This approach remains widely used in the empirical and theoretical literature, despite concerns over its implicit assumptions, such as that the physiological effects of chronic and acute exposure are equivalent. A key limitation in prior empirical studies assessing rate summation is in the assessment of single endpoints, and not in the responses that lead to those endpoints, thereby concealing the drivers behind discrepancies between predicted and observed responses. Consequently, in Chapters 2, 3 and 4, I developed novel computer vision tools for measuring physiological responses with a granular temporal resolution in developing animals. These tools enabled me to assess how embryos of the great pond snail, *Lymnaea stagnalis*,

develop differently under constant and diurnally fluctuating temperatures, and the extent to which TPCs based on performance in CTs can describe embryonic responses to FTs, not just for single endpoints but throughout embryonic development. This thesis demonstrated that 1) embryonic thermal performance does not remain consistent throughout development in response to chronic exposure, but exhibits significant reductions from early to late development (Chapter 4), 2) that chronic and acute exposure do not have the same physiological effects, evidenced by marked differences in growth under chronic and acute exposure at the same temperatures (Chapter 5), 3) developmental pathways vary based on prior thermal conditions, as shown by differing growth trajectories under constant versus fluctuating temperatures (Chapter 5), and 4) that rate summation poorly predicts developmental endpoints in embryos developing under diurnally fluctuating temperatures, but it may have application to specific developmental periods given the accurate predictions observed for early embryonic periods in this thesis (Chapter 5). Together, these findings demonstrate that responses to constant temperatures are not good predictors of how animals develop under diurnally fluctuating temperatures. Further empirical data over a wider range of thermal conditions will enable greater understanding of the drivers behind responses to fluctuating temperatures, and combining digital imaging with automated bioimage analyses would be an effective approach to accomplish this objective.

Table of Contents

Figure legends

Figure 1.1. Relationship between temperature and developmental rate for *Lymnaea stagnalis* embryos incubated to a range of constant temperatures (A), the solid line represents a thermal performance curve fit using the function outlined in Rezende et al., (2019) with key thermal performance parameters highlighted. The effect of Jensen's inequality on this thermal performance curve is shown for different amplitudes of thermal fluctuation (B), where performance under fluctuating temperatures is calculated as an arithmetic average of performance at the two extremes (i.e min and max) of a given amplitude at each mean temperature (Denny 2017).

Figure 1.2. A thermal performance curve for developmental rate in embryos of *Lymnaea stagnalis* incubated to chronic temperatures (A) is applied to a constant (red lines) and a diel fluctuating temperature regime (black lines) (B) to predict incremental developmental rates (*r*) (C), where predicted time to hatch is quantified as the point at which the cumulative sum of *r* equals one, indicated by the dashed lines (D). Mean predicted developmental rate ($\bar{r}$) as a result of rate summation of these temperature regimes is shown in (A).

Figure 2.1. HeartCV processing pipeline, shown here for hippo stage *R. balthica*, where a given video (A) is down-sampled (B) and used to produce Energy Proxy Traits (Tills et al., 2021, EPTs) then shown in the EPT heatmap (C), where the cardiac region detected via segmentation (D) is

used to filter the video and produce an MPV signal (F). Steps B - D can be supervised *via* the use of a graphical user interface (GUI), where users can scroll through images in the video (upper trackbar) and tune the temporal frequencies (lower trackbar) used to create the EPT heatmap (E). Dynamic visual output is provided comprising segmentation output (left image in GUI) and the EPT heatmap (right image in GUI) (E). Finally, cardiac traits are computed and exported to CSV (abbreviations: Min - Minimum, Max - Maximum, SD - Standard Deviation, RMSSD - Root Mean Square of Successive Differences, IBI - Inter-beat interval) (G). Note that steps a - e can be iterated for scenarios involving multiple time-points.

Figure 2.2. Comparison of heart rate (a) and inter-beat interval measures (b-f; see legend of Fig. 2.1 for abbreviations) quantified using manual counting and HeartCV for the three species studied (*C. intestinalis*: no. juveniles=13, *R. balthica*: no. embryos=30, *P. serratus*: no. embryos=24). Pearson's correlation coefficient and the accompanying p-values are shown in the bottom right of each plot.

Figure 2.3. Relation between temperature and heart rate (HR) (A), mean (B) and standard deviation (C) in inter-beat intervals (IBI) in 2nd ascidian stage *C. intestinalis* (no. juveniles = 8), and the first two principal components of a Principal Components Analysis for HR (*bpm*), minimum (min_b2b), maximum (max_b2b), mean (mean_b2b), standard deviation (sd_b2b) and RMSSD (rmssd) in IBI (D). A bi-plot of these variables is also overlayed with the coordinate space additionally scaled by temperature. HR (E) and IBI measures (F-I) quantified for hippo

stage *R. balthica* exposed to a range of constant temperatures (n=10 for 20°C, 25°C and 30°C). Asterisks indicate degrees of statistical significance determined via one-way ANOVA followed by Tukey's post-hoc comparisons. Non-significant results are indicated by 'ns'. Developmental time-series for HR (J) and mean IBI (K) quantified for three different developmental stages of *P. serratus* in response to chronic elevated temperatures (n=8 per developmental stage for both 15°C and 20°C). Solid lines represent the mean response whilst translucent lines represent individual responses.

Figure 3.1. *Dev-ResNet*: a 3D ResNet network architecture for detecting the onset of developmental events (A). Feature maps at each stage of the network are visualised (A), and these are based on the sample video shown at the start of the schematic (see *'Input Video'* in A). The number of filters at each ResNet layer corresponds to the number of convolutional filters associated with that layer (A). Red text beneath each of the sub-headers denotes the dimensions at a given layer and any changes in resolution achieved by a given layer (A). Layer diagrams for the stem and residual blocks described in (A) are illustrated in (B), where kernel sizes associated with 3D *Conv* and *MaxPool* layers are described by the *Length x Width x Height* notation (e.g. Conv 1x3x3), and stride sizes associated with 3D *Conv* layers are described by *s(x)* notation (e.g. *s1* or *s2*) (B). Residual connections in the 3D ResNet layers are shown using the addition symbol at the end of each block (B). An example developmental event sequence for a single embryo, as predicted by Dev-ResNet, is shown for *L. stagnalis* development at 20°C (C), where event probabilities (z axis) correspond to the confidence of the model in a given event classification. Confusion matrices comparing counts between ground truth and predicted

developmental event labels on the testing dataset for Dev-ResNet (D) and an equivalent 2D-CNN, ResNet-18 (E), trained on the training dataset with selective data augmentation, i.e. class imbalances were removed by applying augmentation to varying degrees depending on a given class' representation in the original data. The following are the number of sample videos associated with each event in the testing data: Pre-Gastrula (n = 280), Gastrula (n = 255), Trochophore (n = 206), Veliger (n = 232), Eye spots (n = 47), Heart beat (n = 127), Crawling (n = 333), Radula (n = 475), Hatch (n = 257) and Dead (n = 126). The axis tick colours assigned to each event are described in (D). The following abbreviations are used throughout the figure: BN: Batch Normalization, ReLU: Rectified Linear Unit, MaxPool: 3D Max Pooling layer, FCN: Fully Connected Network, Conv: 3D Convolutional layer.

Figure 3.2. 2D embeddings derived from the separate application of principal component analysis (A) and UMAP (B) to the output of Dev-ResNet trained using triplet semi-hard loss. Each point in the 2D embeddings represents a single video of a single embryo from the testing dataset (n = 2,339). A sub-plot in both (A) and (B) showcases the relative developmental time associated with each data point, normalised on the basis of time of hatch or death.

Figure 3.3. Timings (mean ± s.d.) of developmental events (A) and time-series event probabilities (B) of embryos of *Lymnaea stagnalis* incubated to a range of chronic temperatures: 15°C (n = 58), 17.5°C (n = 60), 20°C (n = 38), 22.5°C (n = 55), 25°C (n = 38), 27.5°C (n = 59), 30°C (n = 37) and 32.5°C (n = 60). All y axes for (B) have limits of min = 0 and max = 1. Solid lines represent mean whilst shaded regions represent ± 1 s.d. (B). Note that Pre-Gastrula

classifications were excluded in mean event timings as we use this classification for samples associated with development prior to gastrulation, rather than as an event itself.

Figure 4.1. Developmental rate (A) and growth rate (B) from 1st cell division to hatching, with key TPC parameters (thermal optimum: T_{opt}, critical upper limit: CT_{max} and performance at the optimum temperature: P_{max}) highlighted. Box plots represent empirical data, whilst nonlinear solid lines represent the thermal performance curve fit. Thermal performance curve model selection criteria are provided in supplementary material Tables A3 and A4.

Figure 4.2. Developmental event-specific thermal sensitivities of developmental and growth rate in response to chronic incubation temperatures (15°C (*n*=58), 17.5°C (*n*=60), 20°C (*n*=38), 22.5°C (*n*=55), 25°C (*n*=38), 27.5°C (*n*=59), 30°C (*n*=37) and 32.5°C (*n*=60)), where thermal performance in terms of growth or developmental rate is cumulative up until each event (A-H). Points and error bars indicate mean ± sd respectively. See supplementary Table A1 and Fig. A2 for descriptions and example images of each event respectively. Thermal performance curve model selection criteria are provided in supplementary material Tables A3 and A4.

Figure 4.3. Key parameters (thermal optimum: T_{opt}, critical upper limit: CT_{max} and thermal breadth) of developmental event-specific thermal performance curves for developmental (A) and growth (B) rate. Thermal performance curves were constructed from responses to incubation to a range of constant temperatures (15°C (*n*=58), 17.5°C (*n*=60), 20°C (*n*=38), 22.5°C (*n*=55), 25°C (*n*=38), 27.5°C (*n*=59), 30°C (*n*=37) and 32.5°C (*n*=60)).

Figure 5.1. Simulated repeating 24 hr natural thermal regimes, at two mean temperatures: 23.5°C (A) and 17.5°C (B), both with an amplitude of ± 7.7 °C. Water bath temperatures are included to show the offset required to achieve the desired incubation chamber temperatures for a room ambient of 16°C.

Figure 5.2. A TPC fitted to average developmental rate (A, solid black line), constructed from responses of *Lymnaea stagnalis* embryos incubated to constant temperatures (15°C (n=58), 17.5°C (n=60), 20°C (n=38), 22.5°C (n=55), 25°C (n=38), 27.5°C (n=59), 30°C (n=37) and 32.5°C (n=60)) (A), is used to predict hatch time using rate summation for the two fluctuating temperature regimes 17.5 ± 7.7 and 23.5 ± 7.7 °C used in the present study (B). The point at which a cumulative sum of incremental developmental rates (C) equals one (D) is determined to be the predicted time to hatch (C-D). The start of shaded regions in (C) and (D) correspond to the time at which embryos are predicted to have hatched, with colouration corresponding to the two fluctuating temperature treatments outlined in (B). Median-normalised (i.e. physiological time, Wagner et al., 1984) cumulative distribution curves for hatch times observed at constant incubation temperatures were used to fit a Weibull cumulative distribution function (E) for the purpose of introducing variation (samples = 20) around hatch time predicted using rate summation (F). Finally, a TPC fitted to average growth rate (G, solid black line), constructed from responses of *L. stagnalis* embryos incubated to constant temperatures was used to predict mean size at hatch via rate summation, where a cumulative sum (I) of incremental growth rates (H) was performed for the duration of developmental time predicted in (C-D).

Variation in size at hatch is introduced by performing rate summation for the different developmental times predicted in (F). The vertical lines in (I) correspond to the different hatch times predicted in (E-F), and the intersection with the size curves illustrate the size variation introduced.

Figure 5.3. Comparison between predicted responses, quantified from rate summation of TPCs derived from constant temperature incubation, and observed responses for development (A) and growth (B) in *Lymnaea stagnalis* embryos incubated to two fluctuating temperature regimes: 17.7 ± 7.7 (n = 22) and 23.5 ± 7.7 °C (n = 11). The sample size is the same for artificial samples in predicted responses. Statistical significance between predicted and observed data was determined via paired t-test (****: $p < 0.0001$).

Figure 5.4. Comparison between predicted (quantified from rate summation of TPCs derived from constant temperature incubation) and observed developmental responses for development (B-C) and growth (D-E) at key developmental events during *Lymnaea stagnalis* embryonic development (A). Embryos were incubated to two different fluctuating thermal regimes: 17.7 ± 7.7 (n = 2 2) and 23.5 ± 7.7 °C (n = 11). The sample size of predicted samples generated using a Weibull cumulative distribution function (Fig. 5.2E-F) is the same as those in empirical data. Event-specific predictions are quantified via performing rate summation from 1st cell division to a given event, and are not the product of performing rate summation between consecutive events. Statistical significance between predicted and observed data was determined via paired t-test (****: $p < 0.0001$, ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, ns: $p >$

0.05). Absolute deltas between mean predicted and observed values are shown in the subplots (B-E).

Figure 5.5. Size curves (mean ± sd) with developmental periods (i.e. period between consecutive developmental events) overlayed for both fluctuating temperature treatments used in the present study (17.7 ± 7.7 (n = 22) and 23.5 ± 7.7 °C (n = 11)). Mean size curves derived from constant temperature treatments with a comparable average temperatures (17.7°C (n = 60), 20°C (n = 38), 22.5°C (n = 55), 25°C (n = 38)) to the fluctuating temperature regimes of the present study are displayed in the subplot.

Figure 5.6. Individual size curves (A, C: gray lines) throughout embryonic development of *L. stagnalis* exposed to two fluctuating temperature regimes: 17.7 ± 7.7 (A-B) (n = 22) and 23.5 ± 7.7 °C (C-D) (n = 11). The exponential growth phase (A, C, shaded grey region between dash black lines) for embryos in both treatments is used for measuring hourly growth rate (mean ± sd) in response to acute temperature change (B, D), with comparison with predicted growth rates (C, D) based on rate summation of a TPC constructed from chronic temperature exposure (Fig. 5.2G). Both observed and predicted hourly growth rates are normalised for statistical comparison. Horizontal blue lines in (B,D) correspond to temperatures where there are significant differences between observed and predicted growth rates as determined by repeated measures ANOVA (Tukey's HSD, $p < 0.05$).

Figure 6.1. Relationship between temperature and developmental rate for *Lymnaea stagnalis* embryos (A), where the solid line represents a thermal performance curve (TPC) fit to responses observed at constant temperatures, whilst the dashed line represents predicted developmental rate derived from rate summation. Developmental rates observed at two fluctuating temperature regimes (17.7 ±7.7°C and 23.5 ±7.7°C) are also overlaid. Secondly, a 3D landscape illustrating the predicted effect of both amplitude and average temperature on developmental rate (B), where data for amplitudes > 0°C are derived from rate summation of the TPC at an amplitude of 0°C, i.e. constant temperature. In both (A) and (B), rate summation is only performed where the range of temperatures in the fluctuating temperature regime does not exceed the upper and lower critical limits of the TPC.

1 | Introduction

Fluctuations in temperature are inherent within the natural environments most ectotherms inhabit. Such thermal fluctuations can occur across a multitude of temporal scales, from diurnal changes to seasonal shifts (Dillon et al., 2016) and thus even short-lived ectotherms will experience some thermal variability during their lifespan. Because biological processes in ectotherms are temperature dependent and their body temperatures are closely tied to the thermal environment in which they inhabit, the rate of such processes is affected by changes in temperature (Colinet et al., 2015; Morash et al., 2018, 2021). The importance of studying ectotherm responses to fluctuating temperatures (FTs) has long been recognised (Cook 1927; Shelford 1929; Uvarov 1931), but the topic has received increasing attention in the context of climate change (IPCC 2023).

Climate change is altering the natural thermal regimes of almost all ectotherms, and the consequences of such effects are already apparent in many species (Morash et al., 2021). For example, mass mortality events have been observed in critical marine foundation species such as corals (Smale et al., 2019), coral reef fishes (Genin et al., 2020), intertidal mussels (Seuront et al., 2019), salmonid fish (Morash et al., 2021), and others (Coma et al., 2009). A key driver that underpins these mortality events is the incidence of novel, extreme temperatures often in the form of prolonged, abnormally warm periods where the damage accrued over such a period is sufficient to be lethal or increase vulnerability to other stressors, such as disease (Seuront et al., 2019; Smal et al., 2019; Coma et al., 2019), or where acutely high temperatures result in

mortality outright, due to the extent they reach or the rapidity of the warming (Genin et al., 2020; Morash et al., 2021). Though climate change is predicted to continue to affect mean temperatures globally, this will be coupled with changes in thermal variability (Easterling et al., 2000; Meehl & Tebaldi, 2004; IPCC, 2012; Rummukainen, 2012), meaning organisms will experience greater fluctuations in temperature. Yet, the majority of existing research has focused on studying responses to constant temperatures (CTs), which are not representative of the variable thermal environments that ectotherms experience naturally (Colinet et al., 2015; Morash et al., 2018). Thus, it is of increasing importance to study the mechanisms underlying responses to FTs, not only to better understand how ectothermic species operate in their current environment, but also to inform predictions of how they will continue to respond to the challenges of climate change.

Predicting a species' vulnerability to climate change is challenging given that both the thermal environment an organism experiences and its thermal tolerance can vary depending on its ontogenetic stage (Kingsolver et al., 2011; Radchuk et al., 2013). In organisms with complex life cycles, ontogenetic stages with heightened sensitivity can cause life cycle bottlenecks that drive a species' vulnerability to climate change (Thomson et al., 2015; Levy et al., 2015). For many species, early developmental stages constitute the most thermally sensitive life stage (Pandori and Sorte 2018; Dahlke et al,. 2020). Early developmental stages are often immobile (e.g. encapsulated embryos), and thus are unable to behaviourally thermoregulate unlike their adult counterparts to avoid extreme temperatures (Woods et al., 2015). Furthermore, maturation of regulatory functions for coping with temperature (Pörtner and Farrell 2008), pH (Conte 1984)

and oxygen (Spicer 1995) occurs during development, and early developmental stages may therefore be less able to cope with environmental challenge. Organisms also generally accrue energy reserves as they develop and mature, potentially enabling them to better withstand the challenges of environmental stress, such as the energetic costs of heat shock protein (HSP) synthesis (Pandori and Sorte 2018; Pottier et al., 2022). Consequently, increasing importance has been put on resolving the vulnerability of early developmental stages to climate change, both current and future (Pandori and Sorte 2018; Levy et al., 2019; Dahlke et al., 2020). A key component to understanding this vulnerability centres on the responses of early developmental stages to the fluctuating thermal environments that most developing ectotherms inhabit.

Below I provide a brief history of early research into the developmental effects of fluctuating temperatures (1.1). Given the importance of thermal performance curves (TPCs) as an approach in FT research, particularly in research geared towards prediction, I will describe what TPCs are (1.2), discuss their use in research of FTs (1.3) and highlight recent advancements to this research area (1.4). Finally, I will discuss how automated and high-throughput phenomic approaches offer significant promise in understanding developmental responses to the natural thermal regimes that organisms experience *in situ* (1.5).

1.1 | Early findings into the effects of fluctuating thermal regimes

Early research in the 20th century into the effects of FTs on ectotherm development revealed contrasting results relative to CTs. Developmental rate has been a commonly used measure of such effects, and is typically measured as overall time to hatch, metamorphose, or time to

another predetermined event. For example, accelerated developmental rates under FTs relative to CTs were observed at higher mean temperatures in cutworm larvae (1927), codling moths (Shelford 1927) and grasshopper nymphs (Parker 1930). Conversely, decelerated developmental rates under FTs relative to CTs at lower mean temperatures were observed in grain aphis (Headlee 1914) and in fruit flies (Ludwig and Cable 1933). However, Ludwig and Cable (1933) also observed in fruit flies that development could be accelerated under FTs if the mean temperature was above or below the thermal optimum recorded for CTs, but was decelerated at the optimum. Much of this early work was conducted with alternating temperatures, a type of FT regime involving repeated temporary exposure between two constant temperatures, and was particularly focused on understanding whether FTs led to the acceleration or deceleration of development in economically important insect species (Shelford 1927; Ludwig and Cable 1933). The mismatch in developmental rates between CTs and FTs observed in these early studies was at odds with the general consensus at the time, that the relationship between temperature and biological rates was linear for intermediate temperatures (Howe 1967), or for the temperatures organisms typically experience in nature (Uvarov 1931; Ludwig and Cable 1933). If the assumption of linearity held true, then responses to FTs should be approximately equal to those under CTs at the same mean temperature (Uvarov 1931). Given a linear function (e.g. developmental rate as a function of temperature), the average of the function (e.g. average of developmental rates between a given range of temperatures) is equal to the function of the average (e.g. developmental rate at the average temperature) (Denny 2017). However, the same does not hold true of nonlinear functions, primarily as a consequence of Jensen's inequality (Jensen 1906). Jensen's inequality states that for any nonlinear function, the average of the

function $f(x)$ will not be equal to the function of the average $f(X)$ (Ruel and Ayres, 1999). This property has two distinct effects depending on the direction of nonlinearity: 1) in convex curves, $f(x) > f(X)$, and 2) in concave curves, $f(x) < f(X)$ (Jensen 1906)(see Fig. 1.1). In the context of the relationship between temperature and any rate of reaction, this property of nonlinear temperature-rate curves has also been referred to as the Kaufmann effect (Messenger and Flitters 1958; Keen 1978; Worner 1992), after Kaufmann (1932). The effects predicted by Jensen's inequality are evident in the early empirical data of responses to FTs in the above examples, indicating that the relationship between temperature and rate in a given biological trait is nonlinear in ectothermic species (Uvarov 1931; Ludwig and Cable 1933, Messenger and Flitters 1958).

Generally, when developing ectotherms are incubated at a range of constant temperatures, a convex, though sometimes linear, relationship is observed between temperature and rate in a given trait for low to intermediate temperatures, and a concave relationship is observed from intermediate to high temperatures including a peak which is often defined as the optimum temperature (Angiletta 2009). This temperature-rate relationship is often termed a thermal performance curve (TPC) (Angiletta 2009), and it emerged as a powerful tool to link responses to CTs in the laboratory, to FTs in the field (Pradhan 1945; Messenger and Flitters 1958; Keen 1978; Tanigoshi and Logan 1979; Worner 1992). TPCs remain the most widely used approach today for describing the relationship between temperature and any biological rate, including those in ectothermic responses to FTs (reviewed in Sinclair et al., 2016), both for current climatic conditions (Niehaus et al., 2012; Kingsolver et al., 2015; Khelifa et al., 2019; Marshall et

al., 2021) and future climatic scenarios (Deutsch et al., 2008; Kingsolver et al, 2013; Vasseur et al., 2014). Below I will provide an overview of thermal performance curves and their properties, before discussing their use in predicting developmental responses to FTs.

1.2 | Thermal performance curves

Thermal performance curves describe the effect of temperature on a specific biological trait (Huey and Stevenson 1979; Angilletta, 2009), and thus can be used to model the temperature sensitivity of a given species (Colinet et al., 2015). To achieve this, researchers typically use a mathematical function to fit an optimum curve to empirical data (Wagner et al., 1984; Angiletta 2006; Krenek, Berendonk, Petzoldt, 2011). Empirical thermal performance data is generally derived from either static (constant) or ramping assays (Rezende et al., 2014). Static assays expose individuals to a constant stressful temperature until collapse in function of a trait if assessing thermal tolerance, or to a range of constant temperatures until the end of exposure or completion of a developmental process, if assessing temperature dependence (Bubliy and Loeschcke, 2002; Olsson and Uller, 2002). Conversely, ramping assays expose animals to gradually increasing or decreasing temperatures at a given rate, whereby the endpoint is defined as the collapse in the function of a given trait (Rezende et al., 2014). The key thermal performance parameters are often identified using numerical optimization, rather than empirical study (Angilletta 2009). The thermal optimum (T_{opt}), corresponds to the temperature at which performance in a given biological trait is maximal (Angiletta 2009). The upper and lower critical thermal limits (CT_{max} and CT_{min} respectively) generally correspond to temperatures at which some organismal function becomes compromised and may subsequently be lethal if

exposure is prolonged (Cowles and Bogert 1944). Critical thermal limits can be informative measures of thermal performance, as they mark the temperatures at which survival becomes time-limited (Kingsolver et al., 2004). Changes in the thermal breadth of the TPC, i.e the temperature range within which performance is 50% of T_{opt} (Huey and Stevenson 1979), can give us an indication into whether a species is a thermal specialist or generalist. Thermal specialists typically have a narrower thermal breadth but with greater concentration of performance around T_{opt} in comparison with thermal generalists (Huey and Kingsolver 1989; Angilletta et al., 2003; Kingsolver et al., 2004). Horizontal shift in the entire TPC can indicate genetic variation owing to adaptation to specific thermal conditions (Huey and Kingsolver 1989; Kingsolver et al., 2004). Moreover, increases in the height of the curve with expansions in the critical thermal limits are thought to indicate 'winners' as opposed to the opposite for 'losers', whereby the former consists of individuals with both higher optimal performance and wider tolerance limits than conspecifics as a consequence of genetic variation (Kingsolver et al., 2004; Palaima and Spitze 2004). Shifts resulting in a concentration of performance towards cooler or warmer temperatures can also indicate thermal acclimation, via cold or heat hardening respectively (Lagerspetz 2006); thermal acclimation involves a typically reversible change in performance following exposure to a stimulating temperature (Lagerspetz 2006).

Thermal performance curves in ectotherms are typically asymmetric and nonlinear (Angiletta 2009) and thus possess a mathematical property known as Jensen's inequality (Jensen 1906; Denny 2017) (Fig. 1.1). Within the context of thermal performance curves, this property is hypothesized to explain why FTs elicit higher performance below T_{opt} than CTs, and why

fluctuations in temperature will generally result in worse performance at T_{opt}, relative to constant conditions (Fig. 1.1). These effects of Jensen's inequality have been recorded in empirical studies across a variety of traits and species (e.g. Foray et al., 2014; Bartheld et al., 2017; Bernhardt et al., 2018; McCalla et al., 2019; Vajedsamiei et al., 2021). For example, the contraction of a left-skewed curve that we would expect for FT regimes (Fig. 1.1B) has been shown in growth rates of a species of green algae, *Tetraselmis tetrahele* (Bernhardt et al., 2018) and developmental rates of a citrus parasitoid wasp, *Tamarixia radiata* (McCalla et al., 2019). In both of these studies, T_{opt}, CT_{max} and P_{max} (performance at T_{opt}) were all reduced in FTs relative to performance quantified for CTs. This reduction in performance at T_{opt} in FTs has also been found in traits related to fitness and energetic resources in the parasitic wasp, *Venturia canescens* (Foray et al., 2014), swimming speed in helmeted water tadpoles, *Calyptocephalella gayi* (Bartheld et al., 2017), and for growth, feeding and respiration rates in blue mussels, *Mytulis Edulis* (Vajedsamiei et al., 2021). Conversely, improvements in performance at concave areas of the TPC (Fig. 1.1) has been shown in developmental rates of the asian citrus cyllid parasitoid**,** *T. radiata* (McCalla et al. 2019) and similarly in growth, feeding and respiration rates in blue mussels, *Mytulis edulis* (Vajedsamiei et al., 2021). Interestingly, increasing the amplitude in fluctuations over the temperature range within which the TPC is concave improves performance further (Vajedsamiei et al., 2021). Jensen's inequality is a mathematical law, and so one could predict these responses in any nonlinear function (Denny 2017), whether biotic or not. However empirical results such as these demonstrate the importance of considering variability around environmental means, and thus they have important implications for how we predict biological responses to climate change, and the limitations of relying on responses to constant

temperatures.

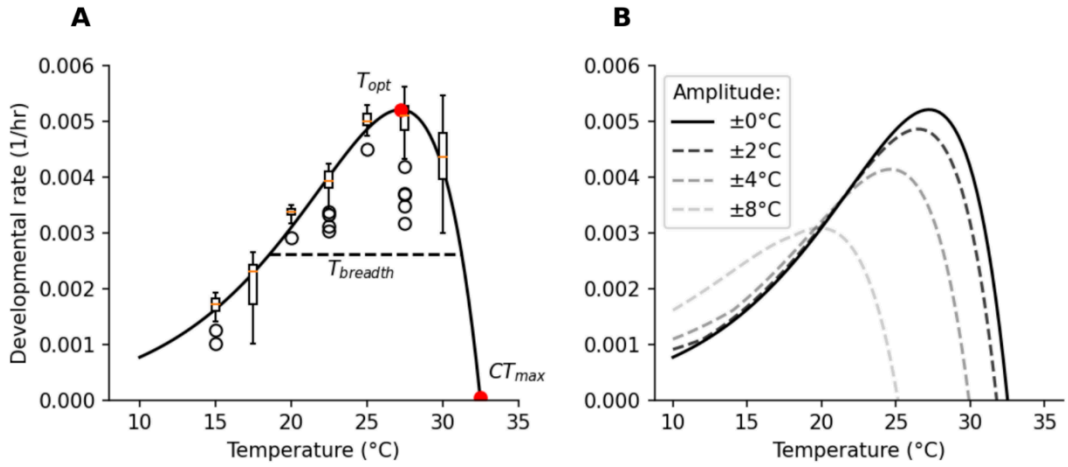

Figure 1.1. Relationship between temperature and developmental rate for *Lymnaea stagnalis* embryos incubated to a range of constant temperatures from 1st cell division to hatch (15°C (n=58), 17.5°C (n=60), 20°C (n=38), 22.5°C (n=55), 25°C (n=38), 27.5°C (n=59), 30°C (n=37) and 32.5°C (n=60)) (A), where the solid line represents a thermal performance curve fit using the function outlined in Rezende et al., (2019) with key thermal performance parameters highlighted (Chapter 4, Fig. 4.1). The effect of Jensen's inequality on this thermal performance curve is shown for different amplitudes in thermal fluctuation (B), where performance under fluctuating temperatures is calculated as an arithmetic average of performance at the two extremes (i.e min and max) of a given amplitude at each mean temperature (Denny 2017).

1.3 | Using thermal performance curves to predict responses to fluctuating temperatures

Researchers have long used thermal performance curves derived from FTs to estimate ectotherm responses to FTs, in both empirical (Keen 1978; Niehaus et al., 2012; Kingsolver et al., 2015; Bernhardt et al., 2018; Khelifa et al., 2019) and theoretical studies (Tanigoshi and Logan 1979; Worner 1992; Georges et al., 2005; Deutsch et al., 2008; Kingsolver et al., 2013;

Vasseur et al., 2014). The most widely used method is to perform an arithmetic sum of incremental rates for the duration of exposure (Pradhan 1945; Tanigoshi and Logan 1979; Worner 1992), where rate at a given time during exposure to an FT regime is derived from a TPC constructed from responses to CT exposure (Tanigoshi and Logan 1979; Worner 1992; Niehaus et al., 2012)(Fig. 1.2). This approach is often termed the rate summation or the rate summing method, and it directly accounts for the effects of nonlinearity inherent within TPCs (Worner 1992; Niehaus et al., 2012; Vasseur et al., 2014; Kingsolver and Woods 2016). For predicting developmental rate (i.e. time from fertilisation to hatch), rate summation is performed iteratively until the cumulative sum equals one (e.g. Tanigoshi and Logan 1979; Worner 1992) (Fig. 1.2D). For other processes with no clear endpoint, such as growth, rate summation is performed for the duration of the experimental exposure period (Niehaus et al., 2012; Kingsolver et al., 2015).

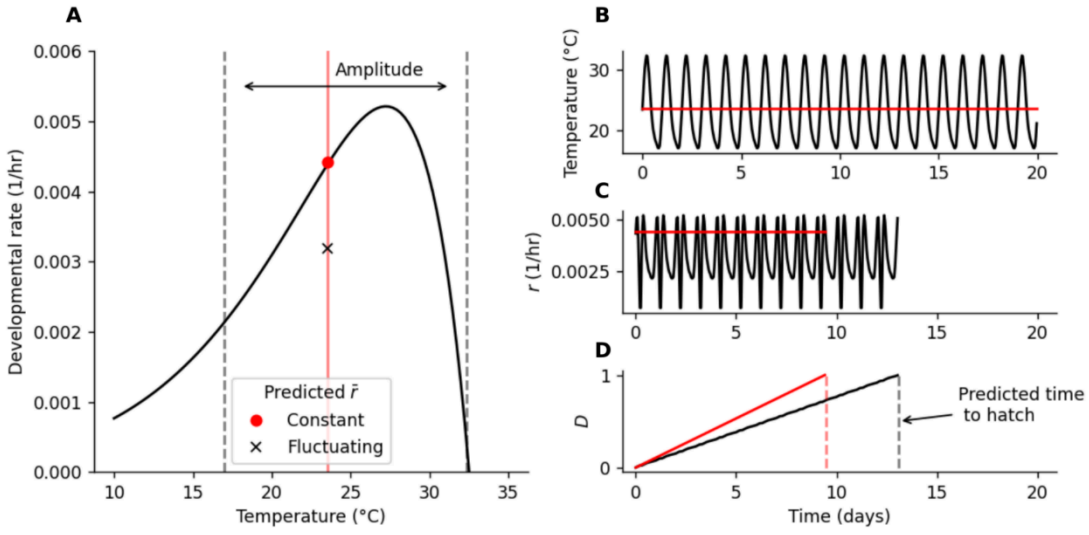

Figure 1.2. A thermal performance curve for developmental rate in embryos of *Lymnaea stagnalis* incubated to chronic temperatures (A, Chapter 4, Fig. 4.1) is applied to a constant (red lines) and a diel fluctuating temperature

regime (black lines) (B, Chapter 5, Fig. 5.1) to predict incremental developmental rates (*r*) (C), where predicted time to hatch is quantified as the point at which the cumulative sum of *r* equals one, indicated by the dashed lines (D). Mean predicted developmental rate ($\bar{r}$) as a result of rate summation of these temperature regimes is shown in (A).

Empirical studies assessing the efficacy of the rate summation approach have often found conflicting or inaccurate results in a developmental context. Early studies largely reported accurate estimation of developmental responses, for example, developmental times were accurately predicted in embryos of the zooplankton, *Chydorus sphaericus* (Keen, 1978), in early life stages of the armyworm, *Pseudaletia unipunctata* (Taylor and Shields 1990) and a range of insect species (Hagstrum and Milliken 1991). However, accurate predictions were only achieved by Taylor and Shields (1990) for FT regimes with a lower average temperature, but not for regimes with a higher average temperature. Similarly, considerable percentage errors were evident between predicted and observed developmental times in the study by Hagstrum and Milliken (1991). Later studies reported largely inaccurate predictions using the rate summation approach, for example in estimating growth and developmental rate for early life stages in the striped marsh frog, *Limnodynastes peronii* (Niehaus et al., 2012), growth rate in tadpoles of the toad, *Bombina orientalis* (Arrighi et al., 2013) and larvae of the tobacco hornworm, *Manduca sexta* (Kingsolver et al., 2015), and in estimating developmental rate in field conditions in *Drosophila embryos* (Khelifa et al., 2019). However, Khelifa et al., (2019) were able to accurately estimate developmental times in *Drosophila* embryos under controlled laboratory conditions with alternating temperature regimes. Conversely, von Schmalensee et al., (2021) accurately estimated developmental rates in the butterfly *Pieres napi*, but for eggs incubated in natural field conditions, and they concluded that *P. napi* developmental rates are relatively unaffected

by fluctuations in temperature, excluding biophysical effects on biological rates. It could therefore be possible that disparities observed in empirical studies are a consequence of active plasticity, which is not accounted for in the rate summation approach given that a single TPC is used to describe trait performance throughout development in response to FTs (Schulte et al., 2011; Niehaus et al., 2012). Whilst rate summation has been used with some success in some empirical studies, there are clearly cases where it is insufficient. These disparities in empirical studies have sparked debate surrounding implicit assumptions of the approach and the use of a single TPC to model responses during long-term exposure or throughout development (Kingsolver and Woods 2016; Sinclair et al., 2016).

There are key implicit assumptions of the rate summation approach that have been hypothesised as driving inaccuracies in empirical studies, namely 1) that chronic and acute thermal exposure have the same physiological effects (Niehaus et al., 2012; Kingsolver et al., 2015; Sinclair et al., 2016), 2) that the relationship between performance and temperature remains constant over the duration of exposure or throughout development (Niehaus et al., 2012; Kingsolver and Woods 2016), and 3) that thermal experience does not alter the response to the current temperature (Kingsolver and Woods 2016). Empirical data disproving each of these assumptions have been recorded as early as the 1930s. For example, Cook (1927) observed that, under FT regimes, larvae of the cutworm, *Porosagrotis orthogonia,* could continue development despite acute exposure to chronically lethal temperatures (assumption 1). Moreover, we might expect thermal performance to change over the course of development on the basis of acclimation or thermal stress (Niehaus et al., 2012)(assumption 2 and 3). For

example, larval growth rates in *Manduca sexta* increased at higher rearing temperatures following short-term (24hr) incubation to FTs as opposed to CTs, potentially indicating beneficial thermal acclimation (Kingsolver et al., 2015). Conversely, in the same study, the thermal optima for growth in *M. sexta* larva declined over the duration of the exposure period under CTs, indicating that performance at higher temperatures becomes limited by thermal stress over time (Kingsolver et al., 2015). Consequently, rate summation of TPCs based on responses to CTs could provide inaccurate predictions of responses to FTs, if thermal stress experienced at extreme CTs is alleviated under comparative FTs, and/or if there are varying degrees of beneficial acclimation in CT versus FT regimes. However, assessing the validity of the latter assumption of the rate summation approach - that the relationship between performance and temperature does not change over the course of development, has been challenging to test in a developmental context, given that the majority of empirical studies in this literature use single endpoints (e.g. developmental time or size at hatch), rather than longitudinal measurement throughout the exposure period (e.g. hourly growth rates). The inaccuracies in predictions found in empirical studies, and concerns over the implicit assumptions of rate summation, have motivated researchers to develop further extensions to the approach.

1.4 | Extensions to the rate summation approach

Recently, a number of advancements to the rate summation approach have been made. Kingsolver and Woods (2016) developed a mechanistic model that integrated time-dependent effects on growth-rate. Time-dependent effects are defined as the effects of duration and pattern of temperature exposure, and can result in either improved or worsened thermal

performance relative to what one would predict using a single TPC alone, i.e. no changes in the shape or breadth of the TPC over the course of exposure (Kingsolver and Woods 2016). In their mechanistic model, Kingsolver and Woods (2016) propose that growth rate (*G*) is determined by proportional control between resource acquisition and the costs of protein production (namely HSPs). Because protein levels present within the system are a function of both temperature (*T*) and time (*t*) (and RNA levels therein), predictions made by this model account for the costs to growth associated with the production of HSPs over time when an animal is exposed to stressful temperatures (Kingsolver and Woods 2016). However, note that resource acquisition in this model is described by ingestion rate multiplied by a conversion efficiency coefficient, where ingestion rate at a temperature *T* is described by a single TPC. Thus, this model cannot account for instances where ingestion rate is itself altered by time-dependent effects. Roitberg and Mangel (2016) developed a model for arthropod growth rate that incorporated a relaxation response from extreme temperatures, where recovery is not instantaneous as it would be when using a rate summation approach, and thus potentially more accurately reflecting the rates of recovery in ectotherms generally (Hazell et al., 2008; Bozinovic et al., 2011; Dittmar et al., 2014). As a consequence of this model, the negative effects of stressful thermal events are exacerbated because ectotherms take longer to recover and repair from the stress induced as a result. Comparing these two models presents areas where their combination might yield new insight. For example, we might predict that the cost associated with thermal stress in a FT regime will be exacerbated; a prolonged recovery from thermal stress would mean that growth, for example, could recover to a lower rate before a subsequent stressful thermal event is experienced. The accumulation of these effects over an exposure period could therefore reduce

the size of an animal beyond what we might predict if we only considered the time-dependent costs of thermal stress, as in the mechanistic model proposed by Kingsolver and Woods (2016).

A key missing component of empirical studies that assess the rate summation approach using FTs, is a lack of longitudinal measurement of the trait of interest for the duration of exposure or development. Without such data, testing assumptions regarding the rate summation approach, and consequently, extensions to the approach as outlined above, are severely limited. The acquisition of longitudinal phenotypic data has become commonplace in other fields involving developing animals, largely driven by technologies associated with the growing field of phenomics (Tills et al., 2018; McCoy et al., 2023). Imaging-based phenomics technologies have enabled longitudinal, high-throughput screening of developing animals, facilitating longitudinal measurement of a wide range of observable phenotypic traits (Spomer et al., 2012; Tills et al., 2018; Gierten et al., 2020; Chapter 2 - Ibbini et al., 2022; McCoy et al., 2023). Thus, for developing animals the use of phenomics technologies appears a natural pairing for attaining a more comprehensive understanding of the drivers behind developmental responses to fluctuating temperatures.

1.5 | Applying phenomics technologies to the assessment of developmental responses to fluctuating temperatures

Phenomics is often defined as the acquisition of high-dimensional phenotypic data on an organism-wide scale (Houle 2010). The history of phenomics dates back to the initial definitions

of the phenome as the “phenotype as a whole” (Davis 1949) and later, “the phenotypic analog of the genome” (Soule 1964). It was not until much later that phenomics gained traction as an approach and saw adoption in a multitude of fields (Bilder et al., 2009; Houle et al., 2010). This was especially helped by advancements in the capabilities of sensor modalities, such as bioimaging, and increases in the affordability and power of computing hardware and associated software pipelines, such as bioimage analysis (Zhou and Wong 2006; Swedlow et al., 2009; Walter et al., 2010; Tills et al., 2018). Such advancements, together with an increase in open source and community driven technologies (Schindelin et al., 2012; Baden et al., 2015; Tills et al., 2018), have substantially reduced the barrier to entry to researchers, and facilitated a far greater adoption of phenomic approaches, including in ecotoxicology (Audira et al., 2020, 2021; Li et al., 2019), crop science (Tardieu et al., 2017; Yang et al., 2020) and the assessment of environmental responses (Tills et al., 2018, 2021). The application of phenomics to these fields has enabled data collection and consequently, hypothesis testing, far greater than what would be capable using manual and/or univariate approaches.

Early development in many species encompasses large phenotypic change in physiology, morphology and behaviour, where manual approaches are ill-suited to capture the breadth of change present within these early life stages. Automated and noninvasive approaches that enable longitudinal measurement of developmental responses therefore have significant potential (e.g. Wong et al., 2010; Spomer et al., 2012; Tills et al., 2018). Indeed, such approaches are commonplace in many phenomics platforms and have enabled responses in early development to be more comprehensively measured. Examples include the identification

of species and stage-specific responses to elevated temperatures in freshwater gastropod embryos (McCoy et al., 2023), the characterisation of growth, movement and cardiac function trajectories to temperature in embryos of *Radix balthica* (Tills et al., 2018), identification of disease phenotypes to toxicants in zebrafish embryos (Capek et al., 2023) and longitudinal cardiac responses of zebrafish embryos to temperature and exposure to different drug compounds (Gierten et al., 2020).

Given the myriad ways in which development can be impacted by temperature, particularly acute exposure, leveraging phenomics technologies has substantial potential in addressing how organisms during the fundamental process of embryonic development respond to fluctuating temperatures. Furthermore, I suggest that the same properties that make aquatic embryos excellent models for phenomics (McCoy et al., 2022), make them excellent models for addressing fundamental questions in thermal biology. Key changes in developmental trajectories could be revealed by using longitudinal measurement, enabling the identification of mechanisms underpinning the discrepancies we see in predicted vs observed responses to fluctuating temperatures. Lastly, assumptions surrounding the rate summation approach can be tested using longitudinal data, but can only be hypothesised with measurement of single endpoints.

1.6 | Concluding remarks and basis of thesis

To date, we know from comparative studies of constant and fluctuating thermal regimes that differences in responses can in some part be explained by the nonlinearity inherent within ectotherm TPCs, which can be accounted for in predictions using rate summation, but that this is not always the sole mechanism driving these differences. It is currently unclear as to what other mechanisms may be at play, particularly because regime design, duration of exposure, traits measured, among other factors, vary considerably between studies, and thus comparing findings between empirical studies can be challenging. Furthermore, the widespread use of CT responses to predict responses to FTs by means of rate summation has received increasing debate, particularly because the implicit assumptions of the approach are at odds with long-established findings of thermal biology research. Consequently, theoretical extensions to this approach have been proposed, though empirical studies to assess the validity of these extensions, and indeed the assumptions of rate summation itself, remain scarce.

Surprisingly, empirical data of how performance in a given trait changes over the course of an exposure period is scarce. Such longitudinal data would provide a clearer picture of how trait performance changes in response to a given regime, whether constant or fluctuating. The scarcity of such data from the literature could be attributed to the technical difficulties in its acquisition, especially for traits where measurable changes occur over longer time periods, such as growth rate. Moreover, manual quantification of changes in trait performance over exposure periods can be very labour intensive, and at times not possible depending on the temporal resolution desired. For example, a higher temporal resolution may be desired for traits where

measurable changes occur over very short periods, such as heart rate or feeding rate. The application of automated, high-throughput phenotyping approaches to this research area therefore has the potential to permit study into questions for which manual approaches would be ill-suited to tackle.

Working in the EmbryoPhenomics research group at University of Plymouth, my research will capitalise on a unique facility for automated phenotyping that permits the acquisition of high dimensional, longitudinal data corresponding to the observable phenotype in early life history stages of aquatic species (Tills et al., 2018). The two primary technologies are the OpenVim microscopy system and the EmbryoCV computer vision software (Tills et al., 2018). OpenVim is a motorised, robotic microscope able to acquire longitudinal bioimaging video, and EmbryoCV is a computer vision software package that extracts a broad range of phenotypic traits from video of embryos of multiple aquatic species (see Tills et al., 2018). These components of *EmbryoPhenomics* can be extended to different species and life history stages, and I present one such extension in Chapter 2 (Ibbini et al., 2022), which describes development and validation of HeartCV, a versatile software for the automated quantification of cardiac traits in transparent animals, that is transferable between species, life history stages and experimental designs. The application of such phenotyping technologies to the study of fluctuating temperatures offers many advantages over manual, univariate approaches, namely that this approach enables automated, longitudinal measurement of the phenotype in response to environmental drivers such as temperature.

1.7 | Thesis aim and outline

The primary aim of this thesis is to measure developmental responses in aquatic embryos to constant temperatures within a TPC framework, and to test the capacity of these measurements for predicting responses to fluctuating temperature regimes. Moreover, this study will utilise automated and high throughput phenotyping technologies to facilitate longitudinal measurement of developmental responses to different thermal regimes. The goal of the combination of these aims with the following methodological approaches is not only to gain greater insight into the drivers behind responses to fluctuating temperatures, but also in how we predict ectotherms will respond to future climate change. Below I will briefly introduce the structure and themes of each chapter of my thesis and subsequently, I will outline key characteristics of my study organism *Lymnaea stagnalis*, the great pond snail, that make it particularly suited to the study of developmental responses to fluctuating temperatures.

In Chapter 2, I develop an automated, noninvasive method for quantifying cardiac traits from video of transparent animals. I demonstrate its capability by validating it against equivalent manual measures in early life history stages of three aquatic species: a freshwater gastropod, *Radix balthica,* a sea squirt, *Ciona intestinalis,* and a prawn, *Paleomon serratus*. Secondly, I apply it to experiments involving different thermal assays with the same three aquatic species to demonstrate the method's functionality and transferability.

In Chapter 3, I develop a 3D convolutional neural network for automated developmental event detection, in order to provide a temporal scaffold of morphological and physiological

developmental events, enabling comparison of the responses of different periods of development to thermal regimes. I test the efficacy of the model in comparison with manually derived event timings in embryonic videos of the pond snail, *Lymnaea stagnalis*. Lastly, I showcase the model's robustness and versatility in experimental contexts by using it to detect developmental events in *L. stagnalis* embryonic development in response to incubation to a range of constant temperatures.

In Chapter 4, I characterise changes in developmental period-specific thermal sensitivity during embryonic development of *L. stagnalis* incubated to a range of constant temperatures (15 - 32.5°C, 2.5°C intervals). This is achieved by quantifying differences in thermal performance curves for both growth and development between key developmental events, and thus permits the assessment of whether the relationship between temperature and performance remains unchanged over the course of development; a key assumption present within fluctuating temperature research.

In Chapter 5, I investigate whether developmental responses to fluctuating temperatures can be predicted on the basis of constant temperature responses in embryonic development of *L. stagnalis*. I compare predicted and observed responses at key developmental events in both growth and development, and also compare constant and fluctuating temperature responses in growth at the same mean temperature, and thus assess the validity of the rate summation approach within the context of embryonic development.

In Chapter 6, I synthesise the core findings of my thesis, including their implications for the study of ectotherm development in fluctuating temperatures. I also discuss gaps in our existing knowledge and highlight avenues for future work.

1.8 | Study species: *Lymnaea stagnalis*

Lymnaea stagnalis is a species of freshwater pulmonate gastropod, and is widely studied across a number of fields, as a bioindicator in ecotoxicology (Amorim et al., 2019), to investigating learning and memory in neuroscience (Fodor et al., 2020), and as a model for developmental biology (Meschcheryakov 1990, Kuroda & Abe 2020). It is found across North America, Europe, Asia (except it's most southern region) and Australia (Hubendick 1951; Atli and Grosell, 2016; Zhang et al., 2018), and it tends to inhabit slow moving or stagnant shallow water bodies, with dense vegetation where it typically grazes on algae, water plants or decaying plant matter (Kuroda and Abe 2020).

L. stagnalis is particularly amenable to laboratory culture, with the largest and longest breeding facility located at Vrije Universiteit in Amsterdam, where this species has been bred continuously for 50 years (Nakadera et al., 2014). Furthermore, *L. stagnalis* is also particularly tractable for developmental studies, owing to it's transparent egg masses and egg capsules (Smirthwaite et al., 2007), enabling the observation of development from first cell division through to hatch (Smirthwaite et al., 2007; Kuroda and Abe 2020; McCoy et al., 2023). *L. stagnalis* egg masses typically contain 100-150 eggs (Nakadera et al., 2014), and eggs can be dissected out of the egg mass with relative ease (Smirthwaite et al., 2007, McCoy et al., 2023),

cultured individually and thus enable high replicate studies to be conducted (e.g. McCoy et al., 2023).

A large number of phenotypic traits are observable in embryonic development of *L. stagnalis*, including growth (Bandow and Weltje 2012; Liu et al., 2022), movement (Byrne et al., 2009) and cardiovascular function (Smirthwaite et al., 2007), making them suitable for the assessment of different stressors on embryonic development, such as temperature (e.g. McCoy et al., 2023). Moreover, the developmental itenerary in terms of events and stages is well characterised (Detlaf and Vazzetsky 1990; Smirthwaite et al., 2007). Thus, the impact of stressors on different periods of development can be studied, since plasticity in this developmental itinerary as a result of a stressor can be compared against pre-established characterisations. Taken together, these attributes of *L. stagnalis* make this species an especially well-suited model for studying the impact of fluctuating temperatures on embryonic development and to answer outstanding questions in the field.

2 | HeartCV: a tool for transferrable, automated measurement of heart rate and heart rate variability in transparent animals

This chapter has been published as:

Ibbini, Z., Spicer, J.I., Truebano, M., Bishop, J. and Tills, O., 2022. HeartCV: a tool for transferrable, automated measurement of heart rate and heart rate variability in transparent animals. Journal of Experimental Biology, 225(19), p.jeb244729.

2.1 | Abstract

Heart function is a key component of whole-organismal physiology. Bioimaging is commonly, but not exclusively, used for quantifying heart function in transparent individuals, including early developmental stages of aquatic animals, many of which are transparent. However, central limitations of many imaging-related methods is a lack of transferability between species, life-history stages and experimental approaches. Furthermore, locating the heart in mobile individuals remains challenging. Here, we present HeartCV: an open source Python package for automated measurement of heart rate and heart rate variability, that integrates automated localization and is transferrable across a wide range of species. We demonstrate the efficacy of HeartCV by comparing its outputs with measurements made manually for a number of very different species with contrasting heart morphologies. Lastly, we demonstrate the applicability of the software to different experimental approaches and to different dataset types, such as those corresponding to longitudinal data.

2.2 | Introduction

Heart rate (HR) and heart rate variability (HRV) are key measures of heart function (Stein et al., 1994; Malik et al., 1996; Acharya et al., 2006), and in many species are proven indicators of cardiovascular health, susceptibility to heart disease and cardiac arrest (Stein et al., 1994; Malik et al., 1996; Norman et al., 2012; Zena et al., 2021). HR and HRV can be measured using a broad range of approaches, including ultrasound, electrocardiography and bioimaging (see Santoso et al., 2020a; Zabihihesari, Khalili, Hilliker and Rezai, 2021). Bioimaging is a common approach for species or developmental stages with a transparent body wall (Brainerd and Hale, 2006; Salmin and Yalcin, 2020; Burggren 2021), and is often preferred over other acquisition technologies because of its transferability between species and life-history stages (Colmorgen and Paul 1995; Peng 2008; Swedlow et al., 2009; Meijering et al., 2016). It is not reliant on a particular experimental setup (Colmorgen and Paul 1995; Schindelin et al., 2012) and is noninvasive unlike other acquisition technologies, such as electrocardiography (Colmorgen and Paul 1995; Zabihihesari, Khalili, Hilliker and Rezai, 2021). Furthermore, advancements in imaging technologies have improved both the image quality and throughput of acquisitions (Shariff et al., 2010; Walter et al., 2010; Swedlow et al., 2009; Salman and Yalcin, 2020). However, the size and magnitude of the image datasets generated using these technologies can extend to millions of images and hundreds of samples (Swedlow et al., 2009; Meijering et al., 2016; Tills et al., 2018). Subsequent manual analysis presents significant data processing bottlenecks and can be limited by the repeatability of analysis and human error (Zhou and Wong 2006; Cardona and Tomancak 2012). Thus, the development of automated computer vision methods to address

these issues is an important endeavor, especially within the broader field of bioimage informatics (Peng 2008; Swedlow et al., 2009; Cardona and Tomancak 2012; Myers 2012).

Many researchers have developed semi- and fully-automated computer vision methods to quantify both HR and HRV (Chan, Lin and Cheng, 2009; Fink et al., 2009; Tills et al., 2018; Gierten et al., 2020; Zabihihesari, Khalili, Hilliker and Rezai, 2021). Generally, these methods are reliant on relatively simple image measures to identify when heart beats occur, such as mean or standard deviation in pixel values (e.g. Chan, Lin and Cheng, 2009; Fink et al., 2009; Tills et al., 2018; Gierten et al., 2020), measures of motion revealed *via* frame subtraction approaches (e.g. Fink et al., 2009; Zabihihesari, Khalili, Hilliker and Rezai, 2021), or M-modes, often termed dynamic kymographs (e.g. Fink et al., 2009; Kurnia et al., 2021). These image measures can be derived from any video produced *via* bioimaging, and therefore should be applicable to any species or life-history stage with visible cardiac activity. However, accurate results typically depend on a localization step to identify the region containing the heart, enabling the removal of non-cardiac noise, such as whole body movement, to generate an informative cardiac signal. Manual localization is used by a number of techniques (Santoso et al., 2020b; Kurnia et al., 2021), but this greatly limits the size and scale of datasets that can be analyzed. Some methods instead omit localization entirely, requiring sample preparation to limit non-cardiac noise, often via anesthetization, immobilization and/or dissection (Chan, Lin and Cheng, 2009; Fink et al., 2009; Hoage et al., 2012; Yozzo et al., 2013; Santoso et al., 2020b). Such sample preparation is time-consuming, often invasive and non-reversible, and generally limits applications of the method to tested species and/or specific experimental designs. A number of methods of

automated localization have been developed to overcome these issues (e.g. Spomer et al., 2012; De Luca et al., 2014; Pylatiuk et al., 2014; Gierten et al., 2020; Zabihihesari, Khalili, Hilliker and Rezai, 2021), but are often species- or life-history stage- specific, with transferrable methods remaining scarce.

Recently developed techniques that integrate power spectral analysis into software-driven localization have proven effective at measuring both complex integrative physiological signals and heart rate (Tills et al., 2018; Gierten et al., 2020). A key advantage of such spectral methods is their transferability between species and life history stages with considerable non-cardiac motion such as muscular contractions and rotational movements (Tills et al., 2018). Cardiac activity is located on the basis that the rhythmic contractions of a heart produce quasi-periodic fluctuations in pixel values between images in videos. Applying power spectral analysis to these fluctuations in pixel values enables identification of the temporal frequencies at which cardiac activity is occurring, and this can therefore be used to identify regions of interest based on the likelihood that they contain cardiac activity (Tills et al., 2018; Gierten et al., 2020). The potential for spectral techniques to support a method capable of capturing both HR and HRV across species, life-history stages and experimental approaches remains untested.

Here, we present *HeartCV,* an open source Python package for measurement of HR and HRV that incorporates a transferrable method for automated localization of cardiac regions. We demonstrate the efficacy of HeartCV by comparison with manual measures for disparate aquatic

species with radically different heart morphologies: a) an ascidian (sea squirt), *Ciona intestinalis*, b) a mollusc (snail), *Radix balthica* and c) a crustacean (prawn), *Palaemon serratus*. *Radix balthica* and *P. serratus* possess superficially globular hearts (Smirthwaite et al., 2007; Dantzler, 1997), whereas *C. intestinalis* has a tubular heart (Davidson 2007), enabling the testing of automated localization of cardiac signals without reliance on a particular cardiac morphology. To evaluate the ability of *HeartCV* to measure HR and HRV responses in different experimental contexts, we used: a) ramping thermal assays in *C. intestinalis*, b) static thermal assays in hippo stage *R. balthica* embryos and c) chronically elevated temperatures at different developmental stages of *P. serratus*.

2.3 | Materials and Method

2.3.1 | Overview of HeartCV software

HeartCV is a Python package for measuring heart rate and interbeat intervals from videos of transparent animals. The software operates using a two-stage workflow involving i) localization to identify the predominant cardiac region in a video and thus extract a mean pixel value (MPV) signal with minimal non-cardiac noise (Fig. 2.1A - F), and ii) peak detection to extract heart beats from this MPV signal (Fig. 2.1F). Subsequently, heart rate (HR) and inter-beat interval (IBI - the timing between individual heart beats) measures are quantified from the MPV signal (Fig. 2.1G). Note here we define inter-beat interval as the timing between individual heart beats. HeartCV can be used in experiments involving either single or multiple time-points (see

https://github.com/EmbryoPhenomics/heartcv/blob/main/examples/example_1.py and https://github.com/EmbryoPhenomics/heartcv/blob/main/examples/example_2.py respectively). Please refer to the documentation for a detailed guide on installation and usage: https://heartcv.readthedocs.io/en/stable/.

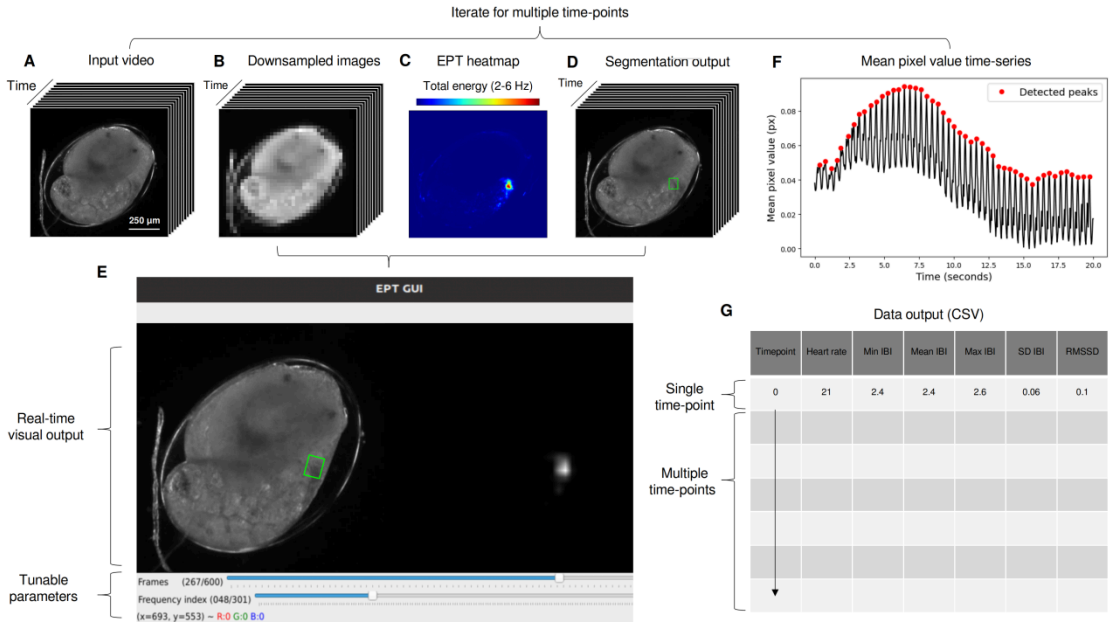

Figure 2.1. HeartCV processing pipeline, shown here for hippo stage *R. balthica*, where a given video (A) is down-sampled (B) and used to produce Energy Proxy Traits (Tills et al., 2021, EPTs) then shown in the EPT heatmap (C), where the cardiac region detected via segmentation (D) is used to filter the video and produce an MPV signal (F). Steps B - D can be supervised *via* the use of a graphical user interface (GUI), where users can scroll through images in the video (upper trackbar) and tune the temporal frequencies (lower trackbar) used to create the EPT heatmap (E). Dynamic visual output is provided comprising segmentation output (left image in GUI) and the EPT heatmap (right image in GUI) (E). Finally, cardiac traits are computed and exported to CSV (abbreviations: Min - Minimum, Max - Maximum, SD - Standard Deviation, RMSSD - Root Mean Square of Successive Differences, IBI - Inter-beat interval) (G). Note that steps a - e can be iterated for scenarios involving multiple time-points.

2.3.2 | Localization

HeartCV initially resizes each image in a given video by a user-defined factor to reduce computational load (Fig. 2.1B). Power spectral analysis is then performed on the resized video to create *Energy Proxy Traits* (EPTs): the amount of energy within different temporal frequencies in the pixel value fluctuations from video of live biological material (Tills et al., 2018, 2021). EPTs are then filtered to only specific, or a range of, temporal frequencies at which cardiac activity is expected. Filtered EPTs are subsequently collapsed into a two-dimensional heatmap by computing the total energy at these specified temporal frequencies (Fig. 2.1C). This heatmap is then processed using the OTSU thresholding algorithm (Otsu 1979), and the largest shape found in the thresholded image using contour detection and filtering is identified as corresponding to the largest region of cardiac activity in the video (Fig. 2.1D). This cardiac region is isolated using a rectangular bounding box and is used in all subsequent steps. Where necessary, or preferred, the steps of localizing heart function (Fig. 2.1B - D) can be supervised through the use of a graphical user interface (GUI), to finely tune parameters to identify the temporal frequencies corresponding to cardiac activity (Fig. 2.1E). Note that the users can only select specific frequencies via the GUI (e.g. 2.6 Hz), whereas this limitation is not present when addressing the underlying functions directly in Python. The final step of the localization stage of HeartCV is the calculation of a mean pixel value (MPV) time-series from the filtered image sequence to enable identification of the timing of individual heat beats (Fig. 2.1F).

2.3.3 | Peak detection

The MPV time-series produced in the localization stage is first up-sampled *via* linear interpolation to emphasize the dominance of peaks and thereby to increase the ability of subsequent analyses to identify the timing of individual heart beats. Automatic multiscale peak detection (AMPD, Scholkmann et al., 2012) is then applied to the processed MPV signal to extract peaks that correspond to heart beats (Fig. 2.1F). These peaks are directly used to quantify beats per minute (bpm) and inter-beat interval measures.

2.3.4 | Experimental approach

Ciona intestinalis

Adult individual *Ciona intestinalis* were collected by hand from Sutton Harbour Marina in Plymouth, UK (50.22 N, 4.08 W). They were used to found a laboratory culture using the methods of Sato et al. (2012) via *in vitro* fertilization. Juveniles were maintained in aquaria containing native sea water (NSW) (vol. = 5 L, T = 15 °C, S = 34) and were later transferred to larger aquaria (vol. = 7 L) at approx. 10 weeks post-fertilization. Full water changes were carried out weekly and individuals were fed *Isochrysis galbana* and *Chaetoceros gracilis ad libitum.*

Initially, 1st ascidian stage juveniles (n = 13, ~2 weeks post-fertilization) were imaged (720 x 540 px, 25 frames per second (fps), 8 bit depth) to produce videos (duration = 10 sec.) to validate against manual measures. Imaging was carried out using a Sony Nex-5N camera (Sony, Tokyo, Japan) attached to a Euromex RZ microscope. These videos were then processed using a

supervised approach (Fig. 2.1E, see https://github.com/EmbryoPhenomics/heartcv/blob/main/examples/example_1.py) to produce the validation dataset. To address a hardware issue with the camera resulting in frame doubling that effectively halved the number of usable images, we smoothed the MPV signals produced by HeartCV using the LOWESS algorithm with fraction set to 0.015 (Cleveland 1979).

To assess the ability of HeartCV to quantify responses to a rapid ramping thermal challenge, 2nd ascidian stage juveniles (n = 8, approx. 8 months post-fertilization) were imaged (QImaging R3 Retiga camera, QImaging, Birmingham, UK - 480x364 px, 20 fps, 16 bit depth, 10x magnification, attached to a Leica M205C microscope, Leica, Wetzlar, Germany) continuously throughout ramping assays (30 minutes at a ramping rate of 0.4°C min^{-1} from 16°C to 29°C). Individuals were transferred from aquaria to Petri dishes containing NSW and placed on a heated glass table (T-Glass controlled *via* a OkoLab H401-T, OkoLab, Naples, Italy) positioned over a dissecting microscope with darkfield lighting (Leica CLS 150 LED - Wetzlar, Germany). Temperature was monitored using a thermocouple secured in the Petri dish containing the individual and this was logged to the OkoLab H401-T controller. MicroManager (Edelstein et al., 2010) was used to acquire continuous video in the form of OME TIFF stacks that were converted to 8-bit video to be compatible with HeartCV using a custom Python script (https://github.com/EmbryoPhenomics/heartcv/blob/main/utils/conversion_16bit_to_8bit.py). HR and IBI measures were then quantified using an iterative supervised approach, where each video was processed at intervals of 20 seconds (Fig. 2.1, https://github.com/EmbryoPhenomics/heartcv/blob/main/examples/example_2.py). An iterative approach with intervals was required here because the cardiac frequencies changed

during the course of the exposure trials, and so processing the footage in small intervals was required for accurate measurement of HR and IBI measures.

Radix balthica

Embryos of *R. balthica* (n=144) were imaged (x200 mag, 750 x 750 px, 20 fps, 16 bit depth, subsequently converted to 8-bit depth using a custom Python script: https://github.com/EmbryoPhenomics/heartcv/blob/main/utils/conversion_16bit_to_8bit.py) at chronic temperatures of 20°C (n = 42), 25°C (n = 44) and 30°C (n = 32), for 30 seconds every hour from the first cell division until hatching. This was achieved using the OpenVIM system: open source video microscopes for high-throughput time-lapse imaging of developing embryos that incorporate a robotic XY stage (see Tills et al., 2018). Embryos were held in 96-well microtiter plates housed within jacket incubation chambers maintained at either T = 20°C, 25°C or 30°C, located within each OpenVIM. This video dataset was subsampled to extract only videos of hippo stage embryos (20°C (n = 10), 25°C (n = 10) and 30°C (n = 10)). HR and IBI measures were then quantified using the supervised approach to produce the experimental dataset (Fig. 2.1E). One-way ANOVA followed by Tukey's post-hoc comparisons was used to test for differences in treatment responses for each of the cardiac measures produced via HeartCV. This was conducted in R v3.6.3 (R Core Team, 2022).

Finally, videos (n = 28) from this subsample were trimmed to the first 10 seconds for manual validation. Two replicates from the 30°C treatment were excluded because whilst usable MPV signals could be produced via HeartCV, their heart rate was too high to perform reliable manual measurement at the frame rate at which the videos were captured. Capturing video at higher

frame-rates would help overcome this issue and so we recommend that users capture video at a frame-rate at which the cardiac cycle can be observed clearly.

Palaemon serratus

Three gravid *P. serratus* females were collected using a hand-held net from tidepools at Jennycliff Bay in Plymouth, UK (50.20 N, 4.07 W). They were maintained in an outdoor tank (vol. = 20L, S = 30-40) for 9 days before being transferred to laboratory conditions, where they were kept at T = ~15**°**C and S = ~35. Individuals were fed marine pellet (New Era Aquaculture™) *ad libitum* throughout, and supplemented with locally sourced seaweeds in the laboratory (*Fucus serratus, Ulva lactuca, Chondrus crispus, Mastocarpus stellatus, Ceramium spp.*, and *Cladophora spp*). Water changes were carried out every other day.

Embryos at each of three different developmental stages were carefully removed from the females' swimmerets when and as required. developmental stages were defined as follows: 'Early': Stage 5, 'Middle': Stage 6, 'Late': Stage 7, in accordance with Müller, Ammar and Nazari, (2004). Embryos from each developmental stage were acclimated to 15 **°**C throughout development until the treatments commenced. Time-lapse recordings of the three developmental stages of *P. serratus* were acquired (x200 magnification, 750 x 750 px, 25 fps, 16 bit depth, subsequently converted to 8-bit depth using a custom Python script: https://github.com/EmbryoPhenomics/heartcv/blob/main/utils/conversion_16bit_to_8bit.py) using the OpenVIM system (Tills et al., 2018) at chronic temperatures of 15**°**C (n = 8 per developmental stage) and 20**°**C (n = 8 per developmental stage), for 24 seconds every hour for 40 hours. Embryos were held in 96-well microtiter plates housed within jacket incubation

chambers maintained at 15°C or 20°C, located within each OpenVIM. Heart rate and measures of IBI were then quantified using HeartCV through an automated approach for multiple time-points (Fig. 2.1, https://github.com/EmbryoPhenomics/heartcv/blob/main/examples/example_2.py), where the upper and lower frequency limits for filtering EPTs was 2 and 6 Hz respectively. Note that for time-points where automated localization was unsuccessful because of little to no cardiac activity (i.e. < 2 heartbeats per time-point), cardiac measures were set to empty values (i.e. NAN). Finally, for manual validation, videos (n = 24, n = 8 per developmental stage, length = 10 seconds, timepoint = 1) from time-lapse recordings corresponding to 15°C were subsequently used or quantification of HR and IBI measures manually.

2.3.5 | Manual validation

Heart rate and inter-beat interval measures from HeartCV were validated by comparison with measures made *via* manual observation of the same videos. Manual analysis was performed using a custom web application developed in Python primarily using Dash (v1.9.1), Flask (v1.1.2, Grinberg, 2018) and OpenCV (v.4.5.2, Bradski 2000) (Fig. A1). The application was used to manually record the image(s) at which heart beats occur for a given video. For the globular hearts of *P. serratus* (n = 24) and *R. balthica* (n = 28), heart beats were recorded at the point of end diastole in the cardiac cycle. Conversely, for the tubular heart of *C. intestinalis* (n = 13), heart beats were recorded at the point when a contractile wave had completed. For each sample, we validated 10 seconds of footage against the equivalent measures produced by

HeartCV. Formal comparison between manual and automated measures of cardiac traits was carried out using Pearson's correlation coefficient via SciPy (Virtanen et al., 2020).

2.4 | Results and Discussion

2.4.1 | Validation

Concordance between manual- and HeartCV-produced data was high for all cardiac measures in all three species tested, despite possessing very different cardiac morphologies (Fig. 2.2). The localization workflow was effective without restrictive sample preparation, such as immobilization or dissection, meaning that HeartCV has promising applicability to experimental contexts in which movement cannot be restricted. The method does rely on direct visibility of cardiac function and so this is a general limitation of the approach, but it is likely to be a less frequent constraint in early developmental stages.

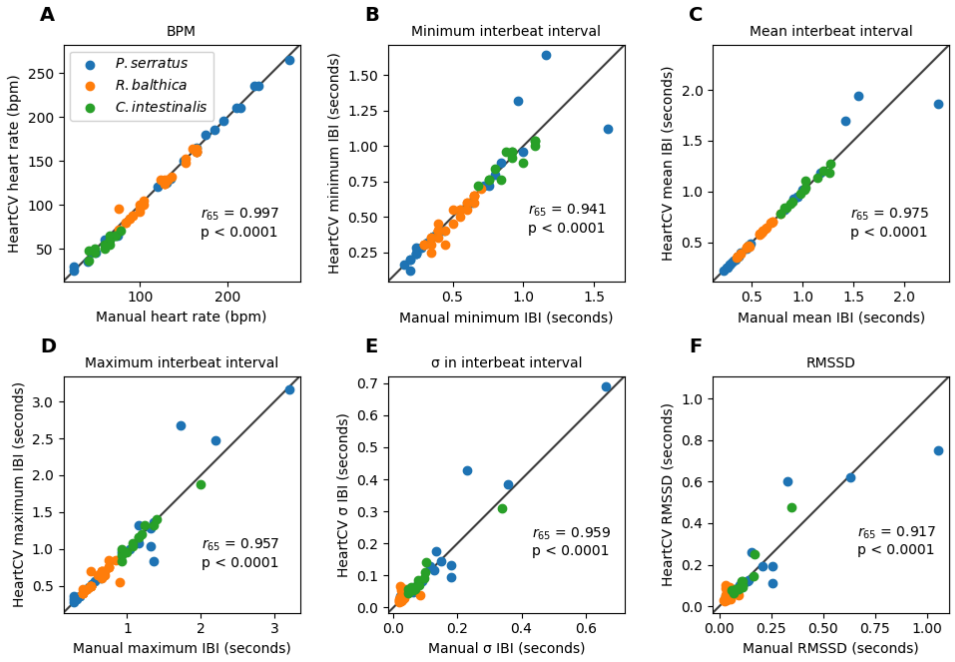

Figure 2.2. Comparison of heart rate (a) and inter-beat interval measures (b-f; see legend of Fig. 2.1 for abbreviations) quantified using manual counting and HeartCV for the three species studied (*C. intestinalis*: no. juveniles=13, *R. balthica*: no. embryos=30, *P. serratus*: no. embryos=24). Pearson's correlation coefficient and the accompanying p-values are shown in the bottom right of each plot.

2.4.2 | Experiment 1: Cardiac response to a rapid thermal challenge in *C. intestinalis*

To test the applicability of HeartCV to dynamic treatments, rapid thermal ramping assays were conducted with 2nd ascidian stage *C. intestinalis* (n = 8). Heart rate and mean IBI exhibited gradual changes with increasing temperature (Fig. 2.3A-B), however, standard deviation in IBI exhibited more erratic trends throughout the assays (Fig. 2.3C). To visualize the integrated cardiac response of *C. intestinalis* encompassing all of the cardiac measures produced by HeartCV, Principal Component Analysis (PCA) was used (Fig. 2.3D). The first two principal

components (PC1, PC2) captured 99.4% of the variance in the cardiac measures. The combinatorial signal quantified via PCA revealed relationships between the various cardiac measures and temperature that were not apparent based on individual comparisons alone. PC1 was primarily driven by heart rate and minimum, mean and maximum IBI (Fig. 2.3D), whereas standard deviation and RMSSD in IBI were most aligned with PC2 (Fig. 2.3D), indicating that these variables are responsible for the breakpoints observed in the coordinate space. The rapid decline in standard deviation in IBI above 26˚C could indicate that these individuals would experience sudden cardiac arrest if the ramping continued to higher temperatures: a consistent short-term reduction in the variation in inter-beat intervals is a key indicator for sudden cardiac arrest in humans (Schechtman et al., 1992; Stein et al., 1994; Tsuji et al., 1994; Hillebrand et al., 2013), but also in mice (Norman et al., 2012) and rainbow trout (Zena et al., 2021). Indeed, other studies measuring thermal performance in *C. intesinalis* have observed that individuals were able to tolerate static exposures of up to 28˚C for 6 h (Serafini et al., 2011) and cardiac arrest was only observed above 32˚C during short term static exposure (<1 h) (Wolf 1932), so the detection of this response at a significantly lower temperature is of particular interest.

2.4.3 | Experiment 2: Cardiac responses to contrasting thermal environments in hippo stage *R. balthica*

To assess the ability of HeartCV to measure cardiac responses to chronic thermal assays, we analyzed video of hippo stage *R. balthica* embryos exposed to three constant temperatures (20, 25 and 30˚C). There were significant effects of temperature on HR ($F_{(2,27)} = 36.29$, $p < 0.0001$),

minimum ($F_{(2,27)}$ = 33.97, p < 0.0001), maximum ($F_{(2,27)}$ = 16.08, p < 0.0001) and mean IBI ($F_{(2,27)}$ = 59.28, p < 0.0001) (Fig. 2.3E-I), but no effect on standard deviation in inter-beat intervals ($F_{(2,27)}$ = 2.957, p = 0.901) (Fig. 2.3I). Whilst heart rate and mean, minimum and maximum IBI responded clearly to increasing temperature, standard deviation in IBI exhibited little change with increasing temperature (Fig. 2.3E-I). These responses are unsurprising given that the thermal optimum (T_{opt}) for respiration in adults has been found to be at 33-38˚C (Schaum et al., 2017), and upper critical thermal limits (CT_{max}) in juveniles were identified at 36-38˚C, using loss in foot attachment as an endpoint (Johansson and Laurila 2016). Therefore, any aberrations in HR and HRV may not be observed until warmer temperatures are reached. Furthermore, development to hatching has been observed at 30˚C in this population (Tills et al., 2018).

2.4.4 | Experiment 3: Cardiac responses to chronic elevated temperatures in three developmental stages of *P. serratus*

Time-series video of three developmental stages of *P. serratus* exposed to chronically elevated temperatures was analyzed using HeartCV. Significant effects of temperature on the cardiac traits measured by HeartCV were evident: embryos at early and intermediate developmental stages exhibited significantly increased HR (young: $F_{(1, 78)}$ = 208.6, p < 0.0001, medium: $F_{(1, 78)}$ = 399.5, p < 0.0001, Fig. 2.3J) and consequently reduced mean IBI at 20˚C relative to 15˚C (young: $F_{(1, 78)}$ = 165.5, p < 0.0001, medium: $F_{(1, 78)}$ = 111.5, p < 0.0001, Fig. 2.3K). Embryos at intermediate developmental stages exhibited more constrained mean IBI at higher temperatures (Fig. 2.3K). Conversely, whilst embryos at late developmental stages exhibited

significantly increased mean HR ($F_{(1, 30)}$ = 55.28, p < 0.0001, Fig. 2.3J) and reduced mean IBI ($F_{(1, 30)}$ = 36.5, p < 0.0001, Fig. 2.3K) early on in the exposure period (0 - 15 h) at higher temperatures, this was followed by a collapse in heart function in most animals (n = 6), evidenced by the marked decrease in mean HR ($F_{(1, 20)}$ = 13.91, p < 0.001) and rapid increase in mean IBI parameters ($F_{(1, 20)}$ = 6.779, p < 0.01) (15 - 25 h) (Fig. 2.3J-K). The gradual increase in mean HR in the remaining timepoints (25 - 40 h) was caused by two remaining embryos whose development continued (Fig. 2.3J).

Together, these responses indicate an increased sensitivity to chronic warmer temperatures in late developmental stages of *P. serratus*. In adult *P. serratus,* loss of the righting reflex has been observed at ~33˚C (Madeira et al., 2015), and so the mortality observed at 20˚C here would suggest heightened embryonic thermal sensitivity. However, it should be noted that Madeira et al., (2014) acclimated adults at 20˚C and the thermal assays were dynamic, in contrast to the acclimation at 15˚C preceding static thermal assays in the present study. The sensitivity of critical thermal limits to methodological context is widely recognized, and a matter of considerable debate (Terblanche et al., 2007; Rezende et al., 2014; Jørgensen and Overgaard 2019). Thus, differences in responses between that of the present study and Madeira et al., (2014) could be attributed to these methodological differences.

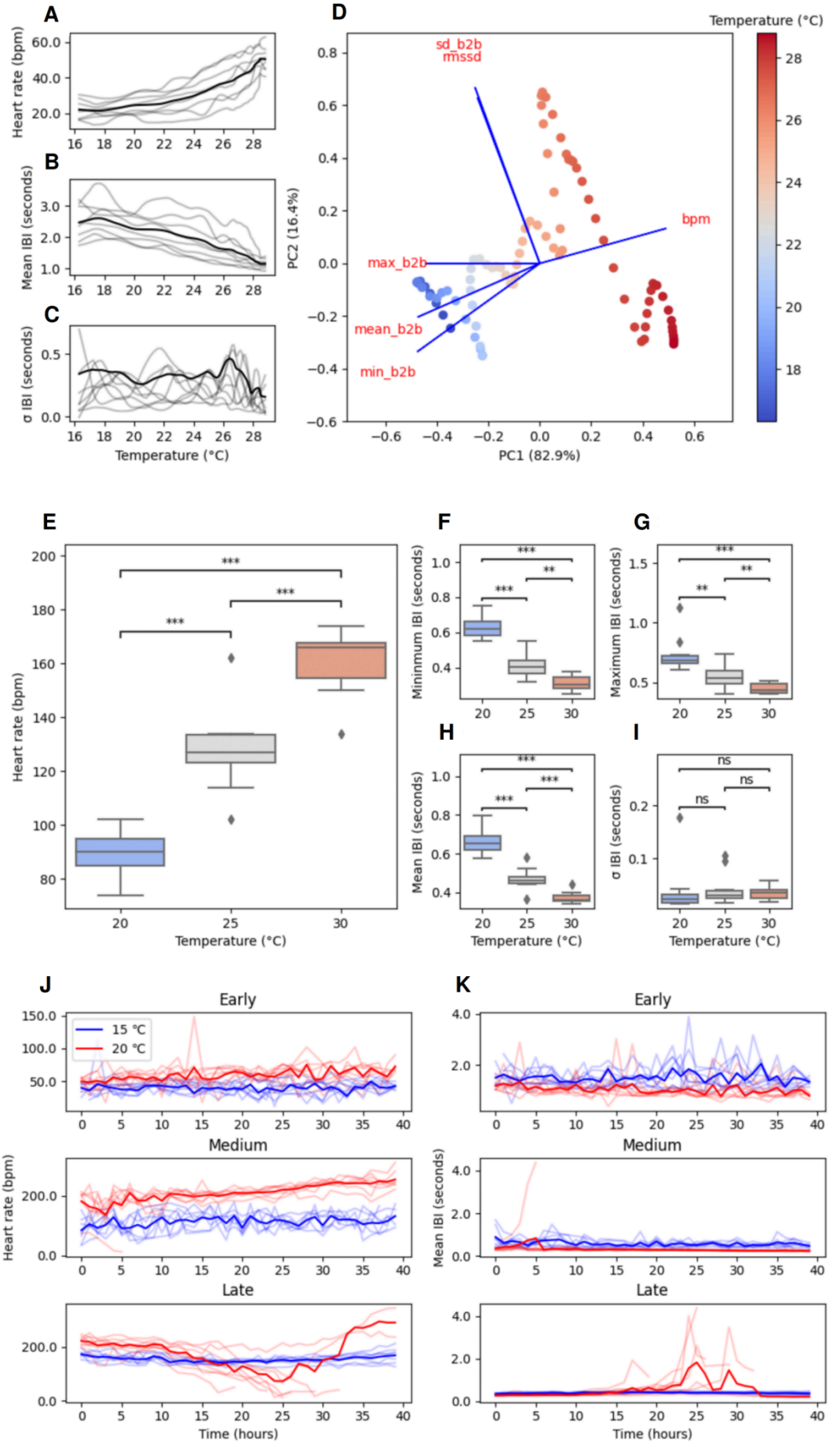
A
B
C
D
E
F
G
H
I
J
K
Heart rate (bpm)
Mean IBI (seconds)
σ IBI (seconds)
Temperature (°C)
PC1 (82.9%)
PC2 (16.4%)
sd_b2b
rmssd
bpm
max_b2b
mean_b2b
min_b2b
Mininmum IBI (seconds)
Maximum IBI (seconds)
ns
Early
Medium
Late
15 °C
20 °C
Time (hours)

Figure 2.3. Relation between temperature and heart rate (HR) (A), mean (B) and standard deviation (C) in inter-beat intervals (IBI) in 2nd ascidian stage *C. intestinalis* (no. juveniles = 8), and the first two principal components of a Principal Components Analysis for HR (*bpm*), minimum (min_b2b), maximum (max_b2b), mean (mean_b2b), standard deviation (sd_b2b) and RMSSD (rmssd) in IBI (D). A bi-plot of these variables is also overlayed with the coordinate space additionally scaled by temperature. HR (E) and IBI measures (F-I) quantified for hippo stage *R. balthica* exposed to a range of constant temperatures (n=10 for 20°C, 25°C and 30°C). Asterisks indicate degrees of statistical significance determined via one-way ANOVA followed by Tukey's post-hoc comparisons. Non-significant results are indicated by 'ns'. Developmental time-series for HR (J) and mean IBI (K) quantified for three different developmental stages of *P. serratus* in response to chronic elevated temperatures (n=8 per developmental stage for both 15°C and 20°C). Solid lines represent the mean response whilst translucent lines represent individual responses.

2.5 | Conclusion

HeartCV is an open-source Python package for noninvasive quantification of cardiac rhythm traits from video of transparent animals and encompasses an effective automated localization technique that is highly transferrable and versatile, making it a powerful tool for experimental biologists.

3 | Dev-ResNet: Automated developmental event detection using deep learning

This chapter has been published as:

Ibbini, Z., Truebano, M., Spicer, J.I., McCoy, J. and Tills, O., 2024. Dev-ResNet: automated developmental event detection using deep learning. Journal of Experimental Biology, 227(10).

3.1 | Abstract

Delineating developmental events is central to experimental research using early life stages, leading to widespread identification of changes in event timing between species and environments. Yet, identifying developmental events is incredibly challenging, limiting scale, reproducibility and throughput of early life stages in experimental biology. We introduce Dev-ResNet, a small and efficient 3D convolutional neural network capable of detecting developmental events with both spatial and temporal components, such as the onset of cardiac function and radula activity. We demonstrate the efficacy of Dev-ResNet using ten diverse functional events throughout the embryonic development of the great pond snail, *Lymnaea stagnalis*. Dev-ResNet was highly effective in detecting the onset of all events, including the identification of thermally-induced decoupling of event timings. Dev-ResNet has broad applicability given the ubiquity of bioimaging in developmental biology, and the transferability of deep learning, and so we provide comprehensive scripts and documentation for applying Dev-ResNet to different biological systems.

3.2 | Introduction

Embryonic development is a fundamental period of life and plays a major role in the evolution of novel phenotypes. Throughout its long history as a focus for biological research (Darwin, 1859; Haeckel 1866; Garstang 1922; Gould 1977), effectively quantifying the continuous process of dynamic temporal and spatial change at the phenotypic level has been a recurring challenge for the field. For over a century, biologists routinely have used developmental event timings as their 'time-stamps'. Events range from the onset of growth of morphological features, through to physiological systems such as the onset of function in cardio- respiratory systems. They have been broadly used in experimental research, including areas such as ecology, ecotoxicology and, most prominently, evolutionary biology. In the latter, developmental events are used to compare the development of different species, as surrogates for chronological and developmental time, and to identify heterochronies, changes in the timing of developmental events between an ancestor and their descendant (Gould 1977; Spicer & Rundle 2006). Heterochronies are a frequent occurrence in living things and posited as key drivers of evolutionary change (McKinney & MacNamara 1991; McNamara and McKinney 2005; Spicer et al., 2011). Developmental event timings have also proven invaluable in comparison of environmental sensitivity between stages of development, and species (Burggren 2021; Rundle & Spicer 2016).

The use and delineation of developmental event timings has attracted controversy (Richardson & Keuck 2022), but they remain widely used with few viable alternatives. A key strength of

developmental event timings is the ability to compare markedly different aspects of development across the animal kingdom, and across a range of environmental conditions. Much of the controversy centers on the precise definition of the characteristics of individual events. No two embryos are identical (Spicer & Rundle 2006), and therefore neither are the appearances of their developmental events. Furthermore, identifying developmental events can be time-consuming and highly subjective, with consequences for the accuracy and reproducibility of the metrics used (Love, 2010). However, advances in methods for analysis of bioimaging datasets are presenting new opportunities. The use of deep learning for bioimage analysis has increased significantly, owing to its potential for generating models capable of solving complex image analysis tasks that are increasingly at the limits of human perception (Nogare et al., 2023).

Deep learning methods typically used for developmental studies involving classification of different embryonic phenotypes involve the use of convolutional neural networks (CNNs). CNNs have been applied to the identification of developmental stages (Liu et al., 2021; Pond et al., 2021), embryonic selection and health measurement for *in-vitro* fertilisation (Chen et al., 2019; Khosravi et al., 2019; Louis et al., 2021), the detection of deformations in embryonic phenotypes (Ishaq et al., 2021; Čapek et al., 2023), and the prediction of developmental time (Jones et al., 2023; Toulany et al., 2023). However, many of these deep learning methods rely on models trained on 2D images (e.g. Liu et al., 2019; Khosravi et al., 2019; Ishaq et al., 2021; Pond et al., 2021; Čapek et al., 2023), and therefore rely solely on morphological features to discern between different embryonic phenotypes of interest. Many phenotypic traits in embryos,

particularly those that are physiological, such as movement and cardiac function, require spatio-temporal data, i.e. videos or sequences of images. Yet, the potential for a model capable of integrating both spatial and temporal information, to perform automated and accurate detection of a diverse range of developmental events throughout development, remains untested.

Here, we introduce Dev-ResNet (https://github.com/EmbryoPhenomics/dev-resnet), a 3D convolutional neural network model capable of detecting both morphological and physiological developmental events throughout the development of embryos, drawing on 3D spatio-temporal information. Dev-ResNet was validated using the highly spatially and functionally dynamic embryonic development of the great pond snail, *Lymnaea stagnalis* (Linnaeus 1758) (Kuroda & Abe 2020). Training of Dev-ResNet was performed on manually labelled videos of 67 embryos (comprising 23,283 videos), to predict the occurrence of ten different developmental events, ranging from gastrulation, to the onset of heart function and crawling. We also tested the applicability of Dev-ResNet to experimental contexts involving different environmental conditions, by imaging the entire development of 405 embryos (comprising 154,082 videos) across a wide range of chronic thermal assays, demonstrating both considerable thermal plasticity in the timing of developmental events, and consistent efficacy of the model.

3.3 | Materials and methods

3.3.1 | Model design

A key point of difference in Dev-ResNet is that instead of using 2D images as inputs and relying solely on morphological differences for delineating between discrete developmental events, Dev-ResNet uses videos of developing embryos converted to a 3D stack of images for input, thus enabling both spatial and temporal features to be used for the detection of developmental events (Fig. 3.1A).

We have opted to use an input video resolution of (length = 120, width = 128, height = 128, channels = 1), and perform a large temporal stride (n = 10) to downsample video to a resolution of (l = 12, w = 128, h = 128, c = 1), i.e. the model only processes one out of every ten video frames, thus increasing computational efficiency whilst retaining global temporal features (Fig. 3.1A). At a typical recording framerate of 30 frames s^{-1}, this temporal stride means that the model makes use of three video frames per second of footage.

The backbone of Dev-ResNet is comprised of a 3D ResNet architecture (Fig. 3.1A), whereby 2D spatial convolutions are coupled with 1D temporal convolutions to produce 3D features at each residual block in the model (Fig. 3.1A-B). This approach to making 3D residual blocks is often termed *Conv2plus1D* (Tran et al., 2018), and has been shown to achieve both increased computational efficiency and comparable, or greater, accuracy when compared to using pure 3D convolutional blocks (Tran et al., 2018).

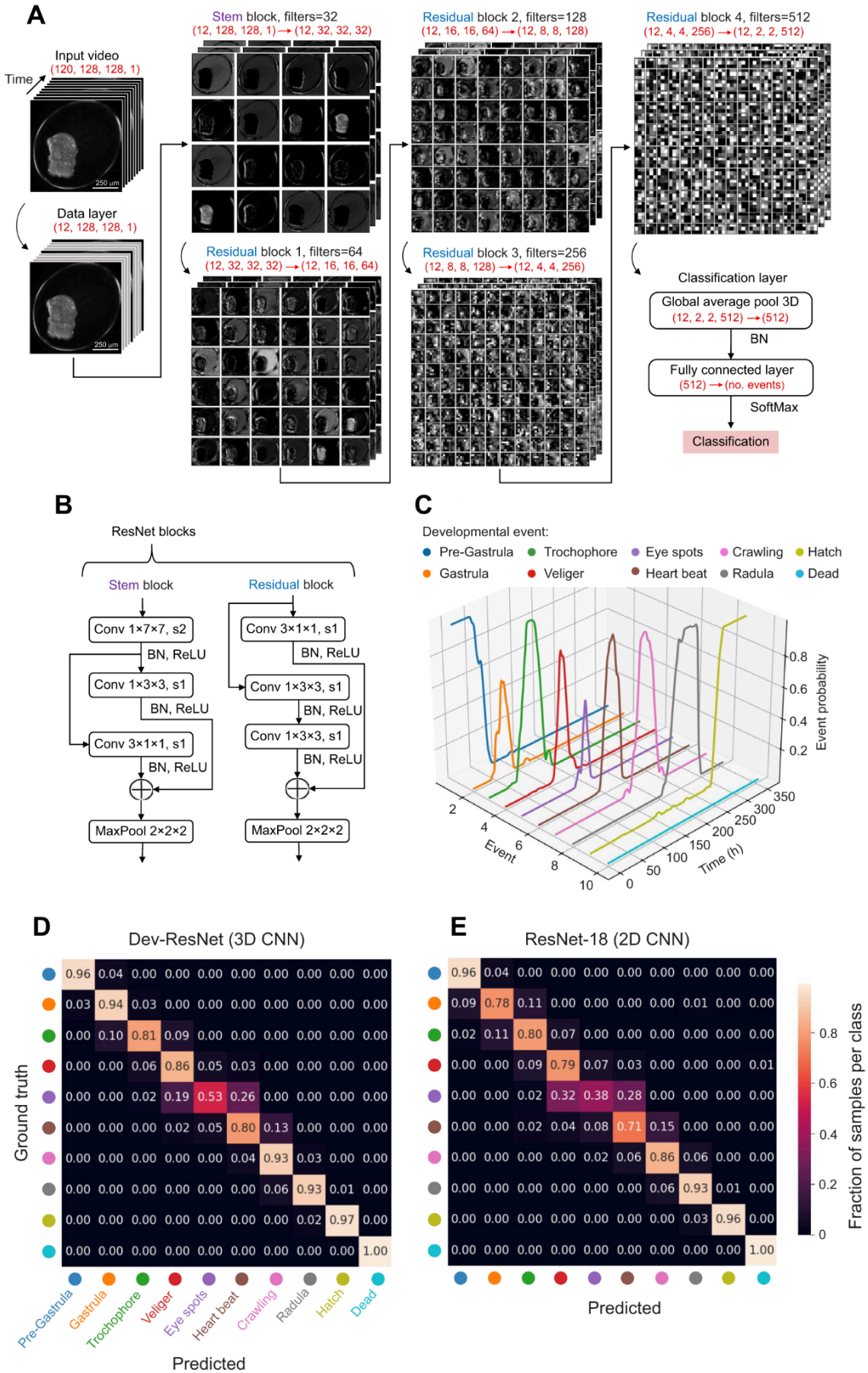
A
Input video
(120, 128, 128, 1)
Time
250 µm
Data layer
(12, 128, 128, 1)
Stem block, filters=32
(12, 128, 128, 1) → (12, 32, 32, 32)
Residual block 1, filters=64
(12, 32, 32, 32) → (12, 16, 16, 64)
Residual block 2, filters=128
(12, 16, 16, 64) → (12, 8, 8, 128)
Residual block 3, filters=256
(12, 8, 8, 128) → (12, 4, 4, 256)
Residual block 4, filters=512
(12, 4, 4, 256) → (12, 2, 2, 512)
Classification layer
Global average pool 3D
(12, 2, 2, 512) → (512)
BN
Fully connected layer
(512) → (no. events)
SoftMax
Classification
B
ResNet blocks
Stem block
Conv 1×7×7, s2
BN, ReLU
Conv 1×3×3, s1
BN, ReLU
Conv 3×1×1, s1
BN, ReLU
MaxPool 2×2×2
Residual block
Conv 3×1×1, s1
BN, ReLU
Conv 1×3×3, s1
BN, ReLU
Conv 1×3×3, s1
BN, ReLU
MaxPool 2×2×2
C
Developmental event:
Pre-Gastrula
Gastrula
Trochophore
Veliger
Eye spots
Heart beat
Crawling
Radula
Hatch
Dead
Event probability
Event
Time (h)
D
Dev-ResNet (3D CNN)
Ground truth
Predicted
E
ResNet-18 (2D CNN)
Predicted
Fraction of samples per class

Figure 3.1. *Dev-ResNet*: a 3D ResNet network architecture for detecting the onset of developmental events (A). Feature maps at each stage of the network are visualised (A), and these are based on the sample video shown at the start of the schematic (see *'Input Video'* in A). The number of filters at each ResNet layer corresponds to the number of convolutional filters associated with that layer (A). Red text beneath each of the sub-headers denotes the dimensions at a given layer and any changes in resolution achieved by a given layer (A). Layer diagrams for the stem and residual blocks described in (A) are illustrated in (B), where kernel sizes associated with 3D *Conv* and *MaxPool* layers are described by the *Length x Width x Height* notation (e.g. Conv 1x3x3), and stride sizes associated with 3D *Conv* layers are described by *s(x)* notation (e.g. *s1* or *s2*) (B). Residual connections in the 3D ResNet layers are shown using the addition symbol at the end of each block (B). An example developmental event sequence for a single embryo, as predicted by Dev-ResNet, is shown for *L. stagnalis* development at 20°C (C), where event probabilities (z axis) correspond to the confidence of the model in a given event classification. Confusion matrices comparing counts between ground truth and predicted developmental event labels on the testing dataset for Dev-ResNet (D) and an equivalent 2D-CNN, ResNet-18 (E), trained on the training dataset with selective data augmentation, i.e. class imbalances were removed by applying augmentation to varying degrees depending on a given class' representation in the original data. The following are the number of sample videos associated with each event in the testing data: Pre-Gastrula (n = 280), Gastrula (n = 255), Trochophore (n = 206), Veliger (n = 232), Eye spots (n = 47), Heart beat (n = 127), Crawling (n = 333), Radula (n = 475), Hatch (n = 257) and Dead (n = 126). The axis tick colours assigned to each event are described in (D). The following abbreviations are used throughout the figure: BN: Batch Normalization, ReLU: Rectified Linear Unit, MaxPool: 3D Max Pooling layer, FCN: Fully Connected Network, Conv: 3D Convolutional layer.

Finally, a 3D global pooling layer and a fully connected layer is used to perform classification (Fig. 3.1A); a classification block that is commonly used in many state-of-the-art 3D classification models (e.g. SlowFast, Feichtenhofer et al., 2019; MoViNet, Kondratyuk et al., 2021).

3.3.2 | Embryo collection

Adult *Lymnaea stagnalis* (n = 25) were sampled from a laboratory stock population and transferred to individual, aquaria (n = 25, vol. = 1.5L, T = 20°C) filled with aerated artificial pond water (APW) (chemical composition: $CaSO_4$ = 120 mg.l^{-1}, $MgSO_4$ = 245 mg.l^{-1}, $NaHCO_3$ = 192 mg.l^{-1}, KCl = 8 mg.l^{-1}) with a 12h:12h light:dark regime. The stock population at University of Plymouth was founded from a culture of the RENILYS strain (donated from IRNA, Rennes, France) established in October 2013 and maintained for multiple generations in large aquaria (n = 3, vol. = 25L, T = 18-22°C, APW). Individuals of the culture used for the present study were fed lettuce *ad libitum* and water changes were carried out 1-2 times per week. Egg masses were removed from aquaria using double layer laminate plastic within 24 hours of deposition, and inspected under low power magnification (10–40×, HM-4, Microtech); egg masses containing embryos that had not developed beyond the 2nd cell division were extricated from the egg mass and transferred to 96-well microtiter plates (Nunc, Microwell, 350 µl per well) for bioimaging.

3.3.3 | Thermal assays and bioimaging

Lymnaea stagnalis embryos (n = 405) were exposed to chronic thermal assays at 15°C (n = 58), 17.5°C (n = 60), 20°C (n = 38), 22.5°C (n = 55), 25°C (n = 38), 27.5°C (n = 59), 30°C (n = 37) and 32.5°C (n=60), and imaged for 20 s (1024x1026 pixels, 30 frames s^{-1}, 8-bit depth, 200x magnification) every hour from first cell division until hatching. This was achieved using the OpenVIM system: an open-source software controlled video-microscope with a robotic X and Y motorized stage for high-throughput time-lapse imaging of developing embryos (Tills et al., 2018). Embryos were held in 96-well microtiter plates housed within jacket incubation

chambers maintained at the treatment-specific constant temperature, located within the OpenVIM. Water levels within microtiter plates were checked 1-2 times per day during imaging and topped up with Milli-Q (Merck, Germany) water as required. The OpenVIM system uses MicroManager (Edelstein et al 2010) to perform image acquisitions which captures videos to individual TIFF files. We developed a python script for concatenating these TIFF files into single timelapse videos for each individual embryo (see https://github.com/EmbryoPhenomics/fast_img_compile/). This facilitated more efficient creation of the training dataset used for training Dev-ResNet. Finally, a pre-trained *Xception* (Chollet 2017) model was used for egg localisation to limit video footage to a 2D bounding box of just the egg. For more information, please see the following GitHub repository: https://github.com/EmbryoPhenomics/egg_detection/.

3.3.4 | Training dataset

The video dataset captured above was subsampled to a smaller dataset for manual annotation (n = 67 individual embryos), with 4-18 replicates per chronic temperature treatment. The hour at which each developmental event occurred was manually recorded by viewing the timelapse video files for each embryo (see Fig. A2 and Table A1 for event images and descriptions respectively). These ground-truth event timings were used to create multi-class annotations, whereby a given video would be labelled as the previous event up until the subsequent event occurred. Finally, we subsampled each timepoint video to every 10^{th} frame of the first 120 video frames (i.e. 4 s of video) and subsequently converted these subsampled videos to GIF files.

The final training dataset size before data augmentation was n = 23,283 videos where each video had resolution: (length = 12, width = 128, height = 128, channels = 1). This dataset was split further into training, validation and testing datasets using an 80:10:10 split, i.e. 80% samples for training and 10% samples for validation and testing each.

The *ImgAug* Python package (https://github.com/aleju/imgaug) was used to apply data augmentation to the training video dataset, where augmentations were deterministic on a sample-specific basis, i.e. the same set of augmentations were applied to each frame in a video sample. Data augmentations applied to the training dataset were: horizontal and vertical flips, amplification (2x) or reduction (0.5x) in pixel values, as well as salt and pepper noise, and Gaussian blur. These data augmentations were used to upsample the dataset size and remove class imbalances between the number of videos associated with a given developmental event. We upsampled all classes to the same number of samples, and the following are the number of augmented videos for classes that were upsampled: pre-gastrula (n = 846), gastrula (n = 1042), trochophore (n = 1170), veliger (n = 1310), eye spots (n = 2652), heart beat (n = 1960), crawling (n = 405), hatch (n = 899) and dead (n = 1835). This type of selective data augmentation facilitated training a far more performant model than non-selective data augmentation (i.e. number of training samples increases but class imbalances remain) (Fig. A4A-C). The full video dataset, including annotations, can be found on Zenodo: https://zenodo.org/record/8214975.

3.3.5 | Training protocol

Training was performed with sparse categorical crossentropy loss and Adam optimizer with a fixed learning rate of 1^{e-6}. We arrived at this learning rate through testing of other learning rates as well as adaptive learning rate schedulers. We found that a fixed learning rate of 1^{e-6} provided the most stable training process as well as the highest testing accuracy. Models were trained for 50 epochs with a batch size of 32 videos.

3.3.6 | Identification of developmental event timings

We identified developmental event timings from probability trajectories associated with each event. Probability trajectories (e.g. Fig. 3.1C) are derived from applying Dev-ResNet to each timepoint video in a given embryo's development and thus produces time-series classifications for each developmental event. We then used two different approaches for identifying developmental events timings from event probability trajectories generated by the final classification layer of Dev-ResNet (Fig. 3.1C). Where event probability trajectories resulted in a peak corresponding to the onset of that event (*Gastrula* through to *Crawling* in Fig. 3.1C), we simply used the time at the maximum probability (i.e. the peak) as the time of onset. Conversely, where an event probability trajectory rapidly increased to a plateau (*Hatch* and *Dead* in Fig. 3.1C), a threshold was also applied but instead the first hour at which Dev-ResNet assigned a probability higher than the threshold was then used as the time of onset. For this latter approach, the following probability thresholds were used: Radula: 0.6, Hatch: 0.4 and Dead: 0.3. Differences in threshold values are due to differences in the upper limits of event probabilities produced by Dev-ResNet for each event (see Fig. 3.1C). These thresholds were

used across all treatments and were identified by comparing predicted event times vs. ground truth data for different threshold values. Thus, we recommend that users of Dev-ResNet perform manual annotations on a sample of their dataset to identify ground-truth timings before computing threshold values.

3.3.7 | Comparison against equivalent 2D architecture

We performed benchmarking comparisons between our 3D ResNet architecture and an equivalent 2D architecture (ResNet-18, He et al., 2016), to establish the importance of temporal information in the detection of developmental events. We trained and tested the ResNet-18 architecture using the same training parameters outlined above for the 3D model (see *Training protocol*), but using only the first frame from each video sample in the training and testing dataset created for the 3D Dev-ResNet model. In doing so, the 2D ResNet model saw the exact same examples during training and testing as the 3D ResNet model, thus enabling us to exclusively test the importance of temporal information in this classification task. Training and testing of both models (2D and 3D) was carried out three times with different random seeds (see https://github.com/EmbryoPhenomics/dev-resnet/train_2d.ipynb), and an average was computed across all iterations for comparisons of top-1 classification accuracy (proportion of classifications where the predicted label exactly matches the ground truth label) and confusion matrices.

3.3.8 | Comparison between different temporal strides and video lengths

To test for the effect of temporal stride and video length on final classification accuracy, we re-trained Dev-ResNet on videos of a range of temporal strides (3, 5 and 10 frame stride, but with the same 4 second length, see https://zenodo.org/records/10702658) and timepoint video lengths (4, 8 and 16 seconds, but with the same 10 frame stride, see https://zenodo.org/records/10719261). The training, validation and testing datasets for each combination of parameters (stride and length) was sourced from the same dataset outlined above, only the number of frames associated with each video sample was different. Selective augmentation was used on all training datasets. We used the exact same training parameters outlined above for the original model (see https://github.com/EmbryoPhenomics/dev-resnet/benchmark_stride_and_length.py), and performed training across three different random seeds for the comparison of top-1 classification accuracy and confusion matrices (Fig. A3D-H). Statistical comparisons of top-1 classification accuracy was performed separately for temporal stride and video length using a one-way ANOVA in R v4.3.2.

3.3.9 | Visualisation of output neurons as 2D embeddings

Leveraging CNNs' ability to distill relevant features from a multi-dimensional input to a single dimensional output, we re-trained Dev-ResNet using triplet semi-hard loss (Schroff et al., 2015) to investigate the relative differences, or similarities, between embryonic videos at different developmental events. Triplet loss trains a model to minimize Euclidean distances between samples of the same class, and maximize distances between samples of different classes. Here,

we replaced the final classification layer of Dev-ResNet with a linear, fully connected layer with L_2 normalization. This fully connected layer had the same number of neurons as the preceding average pooling layer. We subsequently trained this modified model with Adam optimizer with a learning rate of $1e^{-3}$ for 20 epochs. We then applied this model to the testing dataset and processed the output using both principal component analysis (PCA) and the UMAP algorithm (McInnes et al,. 2018) with default parameters, to project the model output to a two-dimensional space for visualisation purposes (see https://github.com/EmbryoPhenomics/dev-resnet/triplet_training_devresnet.ipynb). The following are the key default parameters used for UMAP: $n_{neighbors}$ = 15, $n_{components}$ = 2, *metric* = euclidean, *min_dist* = 0.1, though the full list of parameters can be found here: https://github.com/lmcinnes/umap/blob/master/umap/umap_.py#L1410.

3.4 | Results and Discussion

3.4.1 | Accurate detection of development events with Dev-ResNet

Manual detection of physiological or behavioural developmental events (e.g. onset of cardiac activity or crawling) typically requires real-time spatio-temporal information for confident assessments. Despite this, still images are typically used to train 2D CNNs for classification tasks in developmental studies (Pond et al., 2021; Chen et al., 2019; Zeman et al., 2021; Čapek et al., 2023; Ishaq et al., 2017; Louis et al., 2021).

To test the importance of temporal information for detecting developmental events we compared results with a 2D equivalent (ResNet-18, He et al., 2016) of Dev-ResNet. Dev-ResNet achieved significantly higher classification accuracy than the 2D equivalent (90.3% vs 87.1%) (t_4 = -8.74, $p < 0.001$, for top-1 accuracy). Comparison of confusion matrices (Fig. 3.1D-E), further emphasized the value of a 3D model approach for physiological or behavioural events such as the onset of cardiac activity (ResNet-18: 71%, Dev-ResNet: 80%, Fig. 3.1D-E) and crawling (ResNet-18: 86%, Dev-ResNet: 93%, Fig. 3.1D-E). Whilst morphological characteristics undoubtedly contribute to event classifications, particularly because *L. stagnalis* embryos grow considerably during development, we show that incorporating temporal information via Dev-ResNet significantly enhances classification accuracy. However, adjustments to temporal stride ($F_{(2,6)}$ = 3.175, p = 0.115) and video length ($F_{(2,6)}$ = 2.712, p = 0.145) did not significantly impact classification accuracy of DevResNet (Fig. A3D-H).

3.4.2 | Visualising the continuous process of development

Developmental events mark specific points in a process that is otherwise a continuum of spatial and functional change. However, effectively visualising differences between developmental periods associated with key events remains challenging. Here, visualization of the single dimensional output from Dev-ResNet trained with Triplet loss revealed significant clustering between developmental events, but also a continuum aligned with developmental time (Fig. 3.2).

Videos associated with different developmental events exhibited overlap when visualized using both linear (PCA) (Fig. 3.2A) and non-linear (UMAP) (Fig. 3.2B) dimensionality reductions but mainly for consecutive developmental events, forming a continuum from early to late-stage embryos with clustering within each event (Fig. 3.2). This developmental continuum is further reinforced when points in these 2D embeddings are colored by relative time of occurrence, revealing a continuous gradation of color (Fig. 3.2). Distinct clusters in this continuum are of videos containing hatched (Fig. 3.2B) or dead embryos (Fig. 3.2A-B), which are likely driven by the marked visual differences associated with these events, given that videos of hatchlings contain an empty egg capsule and dead embryos are rapidly colonized by microbes. These data underscore a crucial consideration in the study of developmental events: physiological and morphological change through development is continuous, rather than discrete (Fig. 3.2). Deep learning has enabled the characterisation of the continuous nature of development in other species, such as zebrafish (e.g. Toulany et al., 2023) and thus DevResNet could provide a new objective tool with which to characterise both morphological and physiological development.

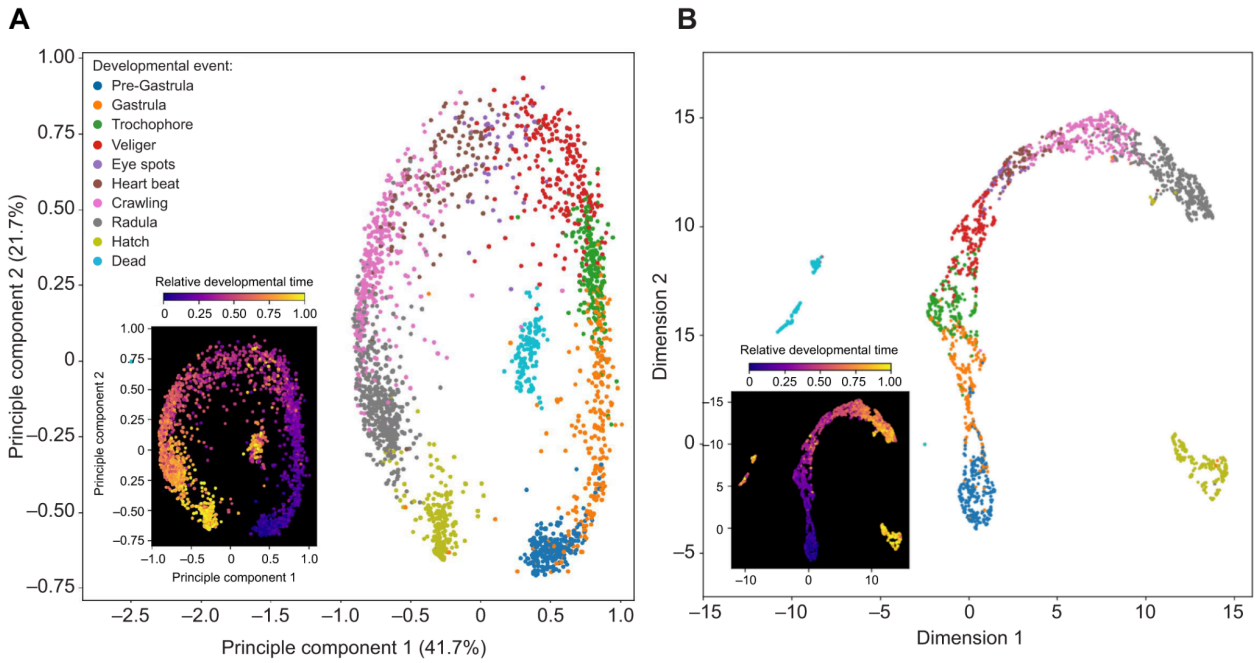

Figure 3.2. 2D embeddings derived from the separate application of principal component analysis (A) and UMAP (B) to the output of Dev-ResNet trained using triplet semi-hard loss. Each point in the 2D embeddings represents a single video of a single embryo from the testing dataset (n = 2,339). A sub-plot in both (A) and (B) showcases the relative developmental time associated with each data point, normalised on the basis of time of hatch or death.

3.4.3 | Dev-ResNet can detect developmental period-specific thermal sensitivities

To test whether a deep learning approach can detect treatment level effects in developmental event timing, we compared the entire development of *L. stagnalis* embryos (n = 405) incubated to a range of chronic temperatures. *Dev-ResNet* successfully automated the analysis of this video dataset and detected developmental period-specific thermal sensitivities: increasing temperature accelerated developmental rate to all developmental events (Fig. 3.3), but comparison between early and late event timings indicated reductions in the optimum temperature (T_{opt}, temperature that developmental rate is highest)(Fig. 3.3A). Such reductions in

T_{opt} for developmental rate highlight heightened thermal sensitivity to upper extreme temperatures in later developmental periods, potentially due to increased costs associated with continued heat shock responses throughout development under chronically elevated temperatures (Tomanek, 2010).

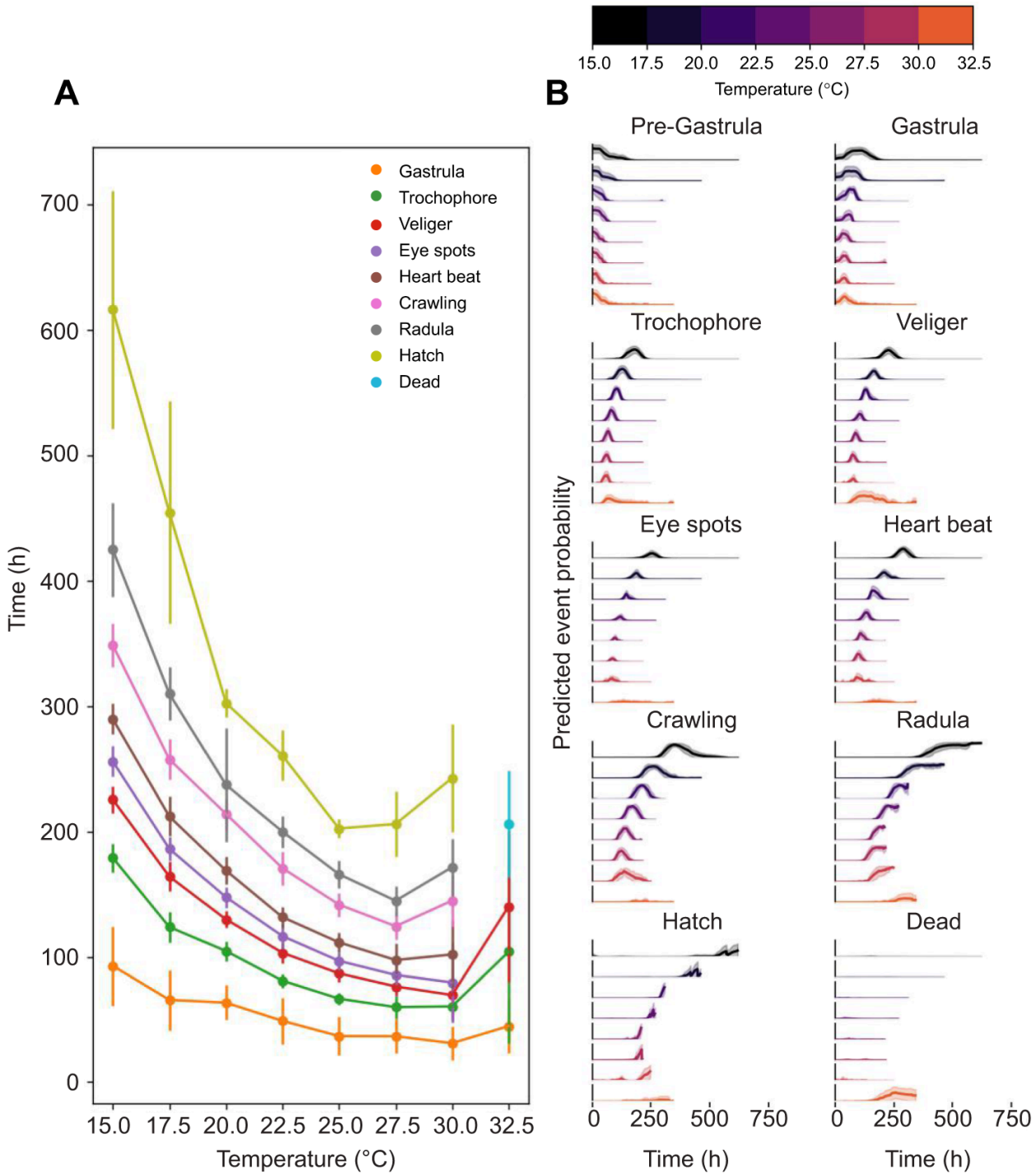

Figure 3.3. Timings (mean ± s.d.) of developmental events (A) and time-series event probabilities (B) of embryos of *Lymnaea stagnalis* incubated to a range of chronic temperatures: 15°C (n = 58), 17.5°C (n = 60), 20°C (n = 38), 22.5°C (n = 55), 25°C (n = 38), 27.5°C (n = 59), 30°C (n = 37) and 32.5°C (n = 60). All y axes for (B) have limits of min = 0 and max = 1. Solid lines represent mean whilst shaded regions represent ± 1 s.d. (B). Note that Pre-Gastrula

classifications were excluded in mean event timings as we use this classification for samples associated with development prior to gastrulation, rather than as an event itself.

3.4.4 | Potential use and limitations

Organismal development, with its inherent physiological diversity and multifaceted time-dependent characteristics has long posed a challenge to researchers who required simplistic approaches to its study. 3D CNNs integrate spatial and temporal features, making them highly relevant to a range of biological characteristics, including morphology and physiology. We found that a single 3D CNN, Dev-ResNet, accurately identified developmental events for which the features of interest are highly different in their appearance and temporal characteristics. Furthermore, Dev-ResNet effectively measured these developmental event timings across a broad range of chronic temperatures, providing a granular resolution in embryonic thermal sensitivity. A key differentiator between Dev-ResNet and other CNNs models for developmental studies (e.g. Jones et al., 2023; Toulany et al., 2023) is the use of discrete functional developmental events, which can change in timing independently of one another and thus enable detection of developmental plasticity. Deep learning approaches such as used here must become increasingly central to the growing field of phenomics, the acquisition of high-dimensional data on an organism wide scale (Houle et al., 2010). Based on this study, deep learning is a powerful solution to the challenge of identifying the timing of events in developing embryos, and this has broad relevance to research across the life and medical sciences.

Dev-ResNet is a comparatively small and efficient model, consisting of just ~5.2 million parameters and ~3.7 GFLOPs respectively (Giga floating point operations). It enables both fast training (320 videos s-1) and inference (541 videos s-1) times on consumer hardware (NVIDIA RTX 3090 GPU), but also achieves high performance using a moderately-sized dataset. These features increase the accessibility to researchers wanting to train models for specific species, developmental periods or events, while reducing the resource cost of both training and inference (Meijering 2020; Hallou et al., 2021; Laine et al., 2021). Generating annotated images or videos for training datasets is a major barrier in deep learning. Thus, a key goal in creating Dev-ResNet was to train an effective model using volumes of data that are attainable for those not working with model systems. Our approach to creating training data involved manual identification of developmental event timings of just 67 embryos (~15% of total replicates imaged in the present study) to produce an accurate, generalisable model capable of accurate inference across a large dataset of 405 embryos.

A potential limitation of Dev-ResNet is its operation on one temporal stride, rather than integrating videos of different temporal strides for classification. An example of this is Google's MoViNet (Kondratyuk et al., 2021), consisting of two 3D-CNNs, each processing either a "fast" or "slow" video input, thereby integrating the different speeds of motion present within video. However, this model type significantly increases computational cost compared to a single 3D-CNN. Another limitation of classification models generally is imbalances between classes. Though we lessened the impact of class imbalances using selective data augmentation, this could not achieve comparable accuracy to developmental events with much more original

samples (Fig. 3.1D). Similar limitations have been observed in other studies (e.g. Liu et al., 2019). An alternative could be to train event-specific models, such as a binary classification model to identify whether a single event occurred or not. However, this approach would linearly reduce training and inference efficiency with the number of event classifications required.

Automated image and video analysis in biology will continue to advance our understanding of complex biological systems (Royer 2023). We suggest that deep learning models like Dev-ResNet are pivotal in broadening deep learning's application beyond traditional model species, thereby increasing its relevance to areas such as evolutionary studies of heterochrony, ecotoxicology, and assessing biological sensitivity to environmental stressors. Extending deep learning approaches to more biological systems will undoubtedly drive scientific advances, and owing to its scalability, contribute to both advancing phenomics and unlocking the phenotyping bottleneck.

4 | Differences in thermal sensitivity during embryonic development of the great pond snail, Lymnaea stagnalis

4.1 | Abstract

Thermal tolerance is a key trait with which organisms cope with stressful temperature conditions, and it has been used extensively to study the geographic distributions of ectothermic species and their vulnerability to climate change. Thermal tolerance can vary between life stages, with embryonic stages often among the most sensitive and vulnerable. Embryonic thermal tolerance is often assessed with static thermal assays, the results of which are typically used to construct thermal performance curves. The use of static thermal assays in developmental studies is widespread, yet data on how thermal tolerance varies during development in response to this assay type remains scarce. Here, we tested for changes in thermal sensitivity at key functional events during embryonic development of the great pond snail, *Lymnaea stagnalis*, incubated to a range of constant temperatures (15, 17.5, 20, 22.5, 25, 27.5, 30 and 32.5°C). We performed timelapse imaging to record embryos every hour from 1st cell division to hatch, with growth and developmental rate measured at key developmental events. We found marked reductions in thermal optima, critical upper limits and thermal breadth, from early through to late developmental events. Though the mechanisms underpinning these responses remain unclear, we highlight that these data are particularly

relevant to the study and prediction of responses to fluctuating thermal environments and future climatic scenarios.

4.2 | Introduction

Climate change is rapidly altering the thermal environments experienced by many ectothermic species (IPCC, 2023), with projected increases in both average temperatures but also the incidence of extreme climatic events (IPCC 2023). Thus, it is crucial to understand how ectotherms will respond to both current and future temperature conditions.

Thermal tolerance is a key trait with which organisms cope with stressful temperature conditions, and it has been used to study the geographic distributions of ectothermic species and their vulnerability to climate change (Deutsch et al., 2008; Sunday et al., 2012; Pinsky et al., 2019). It is widely acknowledged that thermal tolerance can differ between life stages (Storch et al., 2011; Radchuck et al., 2014; Turriago et al., 2015; Truebano et al., 2018; Ruthsatz et al., 2022; Pottier et al., 2022). For some species, this vulnerability is further confounded by early life stages often being sessile and therefore unable to behaviourally thermo-regulate like their adult counterparts. Consequently, thermally-vulnerable developmental stages can lead to a physiological bottleneck, limiting recruitment and thus driving species level impacts in extreme climatic scenarios (Andronikov, 1975). As such, attaining a greater understanding of

developmental thermal tolerance is a prerequisite for accurate prediction of species responses to current and future climate change.

Static thermal assays are frequently used to assess the thermal tolerance of developing ectotherms (Turriago et al., 2015; Truebano et al., 2018; Pottier et al., 2022), especially for whole-organismal responses such as growth (Niehaus et al., 2012; Kingsolver et al., 2015). This approach enables the quantification of thermal performance curves (TPCs) that describe the relationship between temperature and response rate in a given trait, such as metabolic rate or growth rate (Angiletta 2009). Traditionally in developmental studies, TPCs quantified from static assays are then used to infer the effects of fluctuating temperatures (Wagner et al., 1984; Worner 1992; Georges et al., 2005; Niehaus et al., 2012; Khelifa et al., 2019), and thus relate responses in laboratory conditions to those *in situ* and to future climatic scenarios (Deutsch et al., 2008; Vasseur et al., 2014; Sinclair et al., 2016; Khelifa et al., 2019). The ecological relevance of static thermal assays has been heavily debated (Morash et al., 2018), and in particular their use in predictive models for biological sensitivity to climate change (Levy et al., 2015) and fluctuating temperatures (Sinclair et al., 2016), yet they still remain a widely used approach for many aspects of thermal biology. Developmental thermal tolerance to chronic temperatures has been measured across a number of taxa (Turriago et al., 2015; Truebano et al., 2018), though empirical data for how thermal tolerance and performance varies between developmental stages is far more scarce (Pottier et al., 2022). Yet, embryonic development encompasses significant phenotypic change, with a continuum of changing morphology, physiology and behaviour across embryonic stages. However, despite evidence for differences in sensitivity

across embryonic stages, the extent to which thermal performance changes throughout embryonic development in response to chronic incubation temperatures remains untested.

Here, I tested for differences in embryonic stage-specific thermal tolerance in response to a chronic thermal challenge in the freshwater gastropod, *Lymnaea stagnalis*, in two key integrators of performance during the embryonic period i) growth rates and ii) developmental rates. Rates of both growth and development were investigated as, although these traits are often tightly coupled, their differential responses could enable developmental-stage specific responses to be further elucidated. I applied automated, high-throughput imaging coupled with deep learning and computer vision pipelines to characterise stage-specific developmental and growth rates in order to assess how key parameters (thermal optimum, critical thermal limits and thermal breadth) of thermal performance vary throughout the embryonic period of this species.

4.3 | Materials and methods

4.3.1 | Collection of *Lymnaea stagnalis* embryos

Adult *Lymnaea stagnalis* (n = 25) were sampled from a laboratory stock population and transferred to individual, aerated aquaria (n = 25, vol. = 1.5L, T = 20°C) filled with artificial pond water (APW) (chemical composition: $CaSO_4$ = 120 $mg.L^{-1}$, $MgSO_4$ = 245 $mg.L^{-1}$, $NaHCO_3$ = 192 $mg.L^{-1}$, KCl = 8 $mg.L^{-1}$) with a 12 h:12 h light:dark regime. The stock population at University of Plymouth was founded from a culture of the RENILYS strain (donated from the Institute for Agriculture, Food and Environment, Rennes, France) established in October 2013 and maintained for multiple generations (< 10) in large aquaria (n = 3, vol. = 25 L, T = 18-22°C, APW). Individuals of the culture used for the present study were fed lettuce *ad libitum* and water changes were carried out 1-2 times per week. Egg masses were removed from the walls of aquaria using double layer laminate plastic within 24 h of deposition, and inspected under a dissecting microscope (10–40 X, HM-4, Microtech, Wigan, UK); egg masses containing embryos that had not developed beyond the 2^{nd} cell division were extricated from the egg mass and were then transferred to 96-well microtitre plates for bioimaging.

4.3.2 | Thermal assays and bio-imaging

Lymnaea stagnalis embryos (n = 405) were exposed to chronic thermal assays at 15°C (n = 58), 17.5°C (n = 60), 20°C (n = 38), 22.5°C (n = 55), 25°C (n = 38), 27.5°C (n = 59), 30°C (n = 37) and 32.5°C (n = 60), and imaged for 20 s (1024x1026 pixels, 30 frames s-1, 8-bit depth, 200x

magnification) from first cell division until hatching, enabling the measurement of changes in thermal performance throughout embryonic development. This was achieved using the OpenVIM system, an open-source software controlled video-microscope with a robotic X and Y motorized stage for high-throughput time-lapse imaging of developing embryos (Tills et al., 2018). Embryos were held in microtiter plates containing APW (Nunc, Microwell, 96 wells, 350 µL well^{-1}), housed within jacket incubation chambers maintained at one of the constant incubation temperatures, located within the OpenVIM system. Static thermal assays (n = 8) were run in batches on three OpenVIM systems, with eggs from a minimum of three egg masses distributed evenly between microtitre plates assigned to a given thermal assay. Water levels within microtitre plates were checked 1-2 times per day during imaging and topped up with Milli-Q water (Merck, Germany) as required. The OpenVIM systems are controlled using the MicroManager software (v 1.3, Edelstein et al 2010) to perform image acquisitions which capture videos to individual TIFF files. A Python script was developed for concatenating the TIFF files that were acquired into single timelapse videos for each individual embryo, spanning the duration of the experiment for subsequent image analysis (https://www.github.com/EmbryoPhenomics/fast_img_compile/).

4.3.3 | Automated image analysis

Image analysis for (1) localization of eggs, as a precursor to (2) segmentation of embryos to measure size as area and (3) detection of developmental event timings were each automated using deep-learning enabled computer vision, in the form of convolutional neural networks

(CNNs). Training and inference for each image analysis task was performed with an RTX 3090 GPU (NVIDIA, Santa Clara, USA).

Egg localisation

Localisation of *L. stagnalis* eggs was automated through the use of encoder-CNNs, whereby an input image is gradually downsampled through a series of convolutional layers resulting in low resolution features that can be used for computer vision tasks such as image classification or object detection. I re-purposed such encoder-CNNs for localisation of eggs via bounding box regression, whereby the final classification layer of a given CNN is replaced with a fully connected layer with only four neurons, corresponding to the dimensions of a bounding box. Applying this approach to a series of state-of-the-art encoder-CNNs enabled the identification of best performing models for the localisation of *L. stagnalis* eggs (see https://github.com/EmbryoPhenomics/egg_detection/tree/main?tab=readme-ov-file#lymnaea-stagnalis-models).

I created a training dataset of manually defined egg bounding boxes (n = 314) at different developmental periods using the VGG Image Annotator (VIA, https://www.robots.ox.ac.uk/~vgg/software/via/). This training dataset was used to train an initial model (Xception, Chollet 2017) to subsequently assist in producing a far larger training dataset, to support the training of a better performing model for automated localization of eggs from all images from the experiment (n = 174,615). Predicted egg bounding boxes from the initial model were filtered on the basis of their size, given pre-determined size constraints (0.4 - 0.65 mm^2) of *L. stagnalis* eggs and were subsequently visually inspected for exclusion of poor

annotations. These annotations (n = 17,094) were split further into training (n = 13,846), validation (n = 1,709) and testing (n = 1,538) datasets using an 80:10:10 split, i.e. 80% samples for training and 10% samples for validation and testing each. Validation data is used during the training process to provide an unbiased evaluation of the model fit, whereas testing data is only used at the end of the training process to test the final model fit (Hallou et al., 2021). Image augmentations were then applied to the training dataset using the ImgAug library (https://github.com/aleju/imgaug) to reduce overfitting during training. Augmentations applied to the training dataset were: horizontal and vertical flips, amplification (2x) or reduction (0.8x) in pixel values, linear contrast (0.75x to 1.25x), as well as scale transformations (0.75x to 1.25x) and Gaussian blur. These augmentations were applied to all samples in the training dataset to increase samples by a factor of 10 (n = 170,940).

A number of different encoder-CNN models were trained using mean squared error (MSE) loss and Adam optimizer with a fixed learning rate of $1e^{-4}$, for a total of 10 epochs with a batch size of 16 images. For more information regarding training, benchmarking and scripts, please see the following GitHub repository: https://github.com/EmbryoPhenomics/egg_detection.

The best performing model for this particular egg detection task was the EfficientNet V2-S model (Tan and Le 2021), achieving a testing accuracy of 98.1%. This model was then used to perform inference across the entire imaging dataset to produce egg bounding box coordinates for every hour in each embryo's development, enabling videos of embryos to be filtered to just the egg.

Embryo segmentation

I automated embryo segmentation using encoder-decoder CNNs (e.g. UNet, Ronneberger et al., 2015), where an input image is downsampled through several convolutional layers before being upsampled in a mirror-symmetric fashion, thus resulting in an output image with a per-pixel classification; this task is typically termed semantic segmentation. Here, binary semantic segmentation (i.e. background pixels = 0, embryo pixels = 1) was used to generate embryo outlines for computing embryo area. Images used for segmentation were first cropped to just the egg using the model produced for *Egg Localisation* above, and then were subsequently resized with zero-padding to a fixed resolution (w = 256, h = 256, c = 1).

I used VIA manual annotation software to generate embryo outlines for images (n = 314) of embryos at different periods of development, using the drawing tool. Subsequently, I trained a baseline UNet model on these manually annotated images for semi-automated annotation. I developed a simple graphical user interface for visually inspecting and accepting or rejecting embryo outlines generated by the baseline UNet model (https://github.com/EmbryoPhenomics/embryo_segmentation), and this enabled me to create a larger training dataset (n = 3,448). This training dataset was split further into training, validation and testing datasets using an 80:10:10 split, i.e. 80% samples for training and 10% samples for validation and testing each. Image augmentations were then applied to the training dataset using the ImgAug library to reduce overfitting during training.

A number of different encoder-decoder CNN models were trained using binary cross entropy loss and Adam optimizer with a fixed learning rate of $1e^{-4}$, for a total of 25 epochs with a batch size of 32 images. For more information regarding training, benchmarking and scripts, please see the following GitHub repository: https://github.com/EmbryoPhenomics/egg_detection.

The best performing model for the task of segmenting embryos was UNet (Ronnenberger et al., 2015), achieving a testing binary intersection over union (Binary IoU) of 96.2. IoU is a measure of the amount of overlap between ground truth masks and predicted masks, and I use Binary IoU here given that there are only two classes for segmentation predictions: background or embryo. The best performing model, UNet, was then used to perform inference across the entire imaging dataset captured above, producing masks of embryos throughout their development, enabling changes in area to be used to calculate growth rates.

Embryo masks generated using UNet were filtered to pixel classifications above a probability of 0.9, and then the largest contour in these masks was identified as belonging to an embryo. Because the image inputs to UNet were zero-padded images of size 256x256 pixels (width x height) of just the region of interest within the egg bounding box, real-world measures of size in mm^2 were produced via the following formula applied to each embryo contour:

$$Embryo\ area = \frac{mm^2}{px} \times 4 \times \frac{w_1 \times h_1}{w_2 \times h_2} \times Outline\ area$$

Where w_1 and h_1 are the bounding box dimensions (width and height) of the egg detected in the original image, and w_2 and h_2 are the width and height of the images supplied to the UNet model with zero-padding removed. Note that both sets of dimensions are measured in pixels. *Outline area* corresponds to the area in pixels of the mask produced by the embryo outline detected by UNet as computed using the *contourArea* function in OpenCV (Bradski 2000). The outline area is multiplied by four because the original image resolution is 1024x1024 pixels, whilst images supplied to the egg localisation model were 512x512 pixels. Finally, $\frac{mm^2}{px}$ was quantified via first measuring the diameter of an individual *L. stagnalis* egg under a dissecting microscope (10–40×, HM-4, Microtech) against a reference diameter of a single well in a 96-well microtiter plate. Fiji (v1.53) was used to create a pixel per mm^2 scale using an image from the eyepiece of the microscope and thus obtain egg diameter. Given a known egg diameter for a single individual permitted the creation of a pixel per mm^2 scale using Fiji (v1.53) (Schindelin 2012) based on images acquired using the OpenVIM system. A separate $\frac{mm^2}{px}$ scale was quantified for the 30°C treatment due to the use of a different lens, Optem Zoom 70Xl (Optem, Pittsburgh, U.S.A), within the OpenVim system.

Developmental event detection

Hatch time was quantified manually by viewing the time-lapse videos of each embryo and identifying the hour at which 16-cell division and hatching occurred, and this was then used to quantify overall developmental rate (1/h). Finally, the timing of key developmental events was detected through the use of *Dev-ResNet*, a 3D convolutional neural network for automated

detection of developmental events from bioimaging video (Chapter 3 - Ibbini et al., 2024). Dev-ResNet was used to detect the timings of a range of different developmental events, from the onset of gastrulation to the initiation of cardiac function (see Table A1 and Fig. A2 for event descriptions and example images respectively). For a more detailed overview of this approach, please refer to Chapter 3 (Ibbini et al., 2024).

4.3.4 | Statistical analysis

Thermal performance curves (TPCs) were estimated for mean thermal performance data using the rTPC (Padfield et al., 2021) package in R v4.3.1. I fitted separate TPCs for each trait (growth or developmental rate) and corresponding developmental events (see Table A1 and Fig A2). Thermal performance at each development event was quantified as cumulative up until the specific event, i.e. growth or developmental rate from 1st cell division to a given event. For early developmental events (gastrula, trochophore and veliger), CT_{max} was identified on the basis of the thermal performance curve model fit, since embryos incubated at 32.5°C were still able to develop through these developmental periods. However, total mortality was observed at 32.5°C for developmental events beyond the veliger stage, and so CT_{max} was set to this temperature for these events (appearance of eye spots to hatch).

To identify the best model for describing thermal performance at each developmental event, I compared the fits of various mathematical functions using Akaike's information criterion (AIC) (see Johnson and Omland, 2004 and Niehaus et al., 2012). I compared the following functions: Gaussian, Weibull, Thomas et al., (2017) and Rezende et al., (2019) (overview in Padfield et al.,

2021, *rTPC*) (see Table A3 and A4 for comparisons). Two of these functions – the Gaussian and Weibull functions – have been commonly used to theoretically or empirically describe thermal reaction norms (Huey and Kingsolver, 1993; Palaima and Spitze, 2004), whilst the latter functions of Thomas et al., (2017) and Rezende et al., (2019) were selected because their complex structure should provide a better fit to nonlinear data. The model proposed by Rezende et al., (2019) provided the best fit across the most developmental events for both traits (Table A3 and A4), developmental and growth rate, and in cases where this was not the case, the difference in AIC between the best ranking model was minimal (< 2). Thus, this model was used for comparisons of thermal performance metrics between developmental events.

To investigate how metrics of thermal performance change over developmental periods, Pearson's correlation coefficient was quantified separately for three standard metrics of thermal performance curves: temperature at which thermal performance is highest (T_{opt}), upper critical thermal limit (CT_{max}) and the temperature range in which performance is 50% of its maximum (thermal breadth) (Huey and Stevenson 1979; Rezende and Bozinovic 2019). These analyses were performed using the '*pearsonr'* function from the SciPy package in python.

4.4 | Results

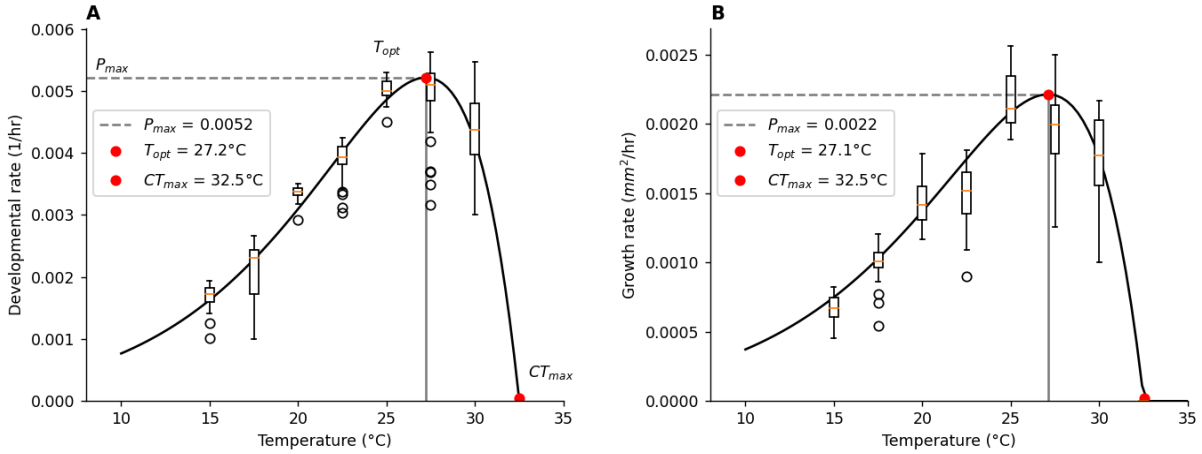

Figure 4.1. Developmental rate (A) and growth rate (B) from 1st cell division to hatching, with key TPC parameters (thermal optimum: T_{opt}, critical upper limit: CT_{max} and performance at the optimum temperature: P_{max}) highlighted. Box plots represent empirical data, whilst nonlinear solid lines represent the thermal performance curve fit. Thermal performance curve model selection criteria are provided in supplementary material Tables A3 and A4.

Rate of development and growth from 1st cell division to hatch was highly temperature dependent (Fig. 4.1), with T_{opt} in both traits estimated at ~27°C whilst CT_{max} was estimated at 32.5°C (Fig. 4.1A-B). Embryos incubated at 32.5°C failed to develop to hatching and thus there was 100 % mortality in this treatment. At this lethal temperature (under chronic conditions), substantial abnormalities were observed in embryos at the veliger stage in the form of large transparent vesicles that deformed identifiable features in the embryos (see Fig. A6).

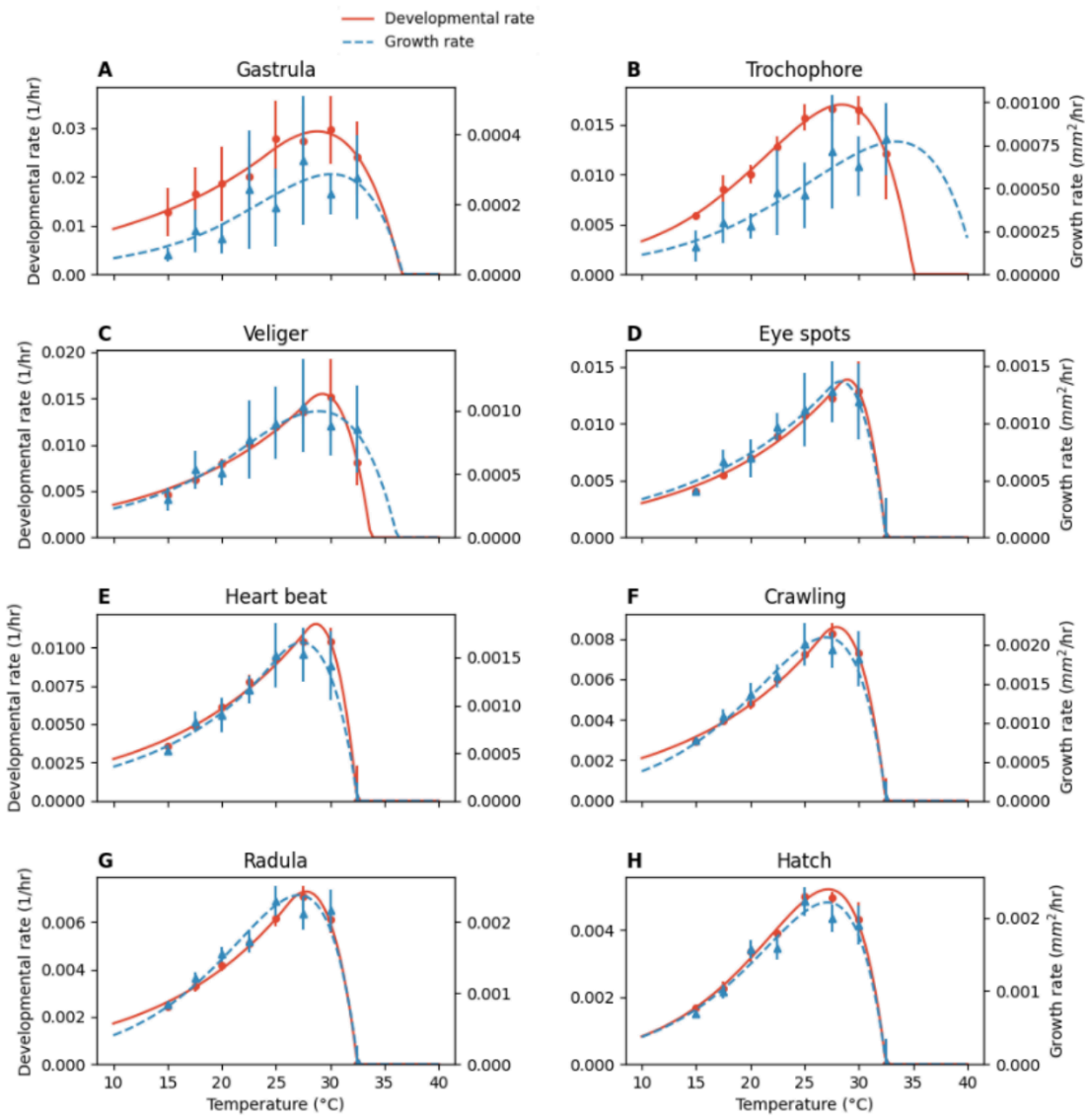

Figure 4.2. Developmental event-specific thermal sensitivities of developmental and growth rate in response to chronic incubation temperatures (15°C (*n*=58), 17.5°C (*n*=60), 20°C (*n*=38), 22.5°C (*n*=55), 25°C (*n*=38), 27.5°C (*n*=59), 30°C (*n*=37) and 32.5°C (*n*=60)), where thermal performance in terms of growth or developmental rate is cumulative up until each event (A-H). Points and error bars indicate mean ± sd respectively. See supplementary Table A1 and Fig. A2 for descriptions and example images of each event respectively. Thermal performance curve model selection criteria are provided in supplementary material Tables A3 and A4.

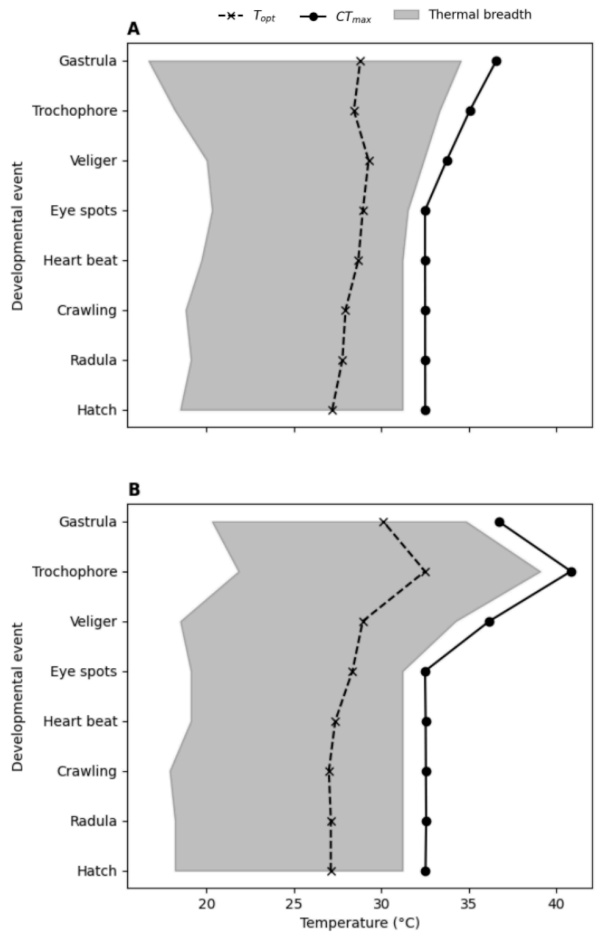

Figure 4.3. Key parameters (thermal optimum: T_{opt}, critical upper limit: CT_{max} and thermal breadth) of developmental event-specific thermal performance curves for developmental (A) and growth (B) rate. Thermal performance curves were constructed from responses to incubation to a range of constant temperatures (15°C (*n*=58), 17.5°C (*n*=60), 20°C (*n*=38), 22.5°C (*n*=55), 25°C (*n*=38), 27.5°C (*n*=59), 30°C (*n*=37) and 32.5°C (*n*=60)).

Comparing TPCs for rates of both growth and development between developmental periods enabled whole-developmental responses to be unpacked further, by identifying periods with differences in performance indicating a decoupling in the responses of growth and functional development (Fig. 4.2 and 4.3). This was evidenced in marked reductions in both T_{opt} (Growth: r_6 = -0.83, $p < 0.05$; Development: r_6 = -0.75, $p < 0.05$) and CT_{max} (Growth: r_6 = -0.77, $p < 0.05$;

Development: r_6 = -0.86, p < 0.05) as development progressed (Fig. 4.3A) (Fig. 4.3B), indicating left-ward shifts in thermal performance through developmental time. Whilst CT_{max} beyond veliger stage is determined by complete mortality observed in the 32.5°C treatment (Fig. 4.2), T_{opt} in both traits continues to decrease until hatching (Fig. 4.3A-B) and this is particularly apparent in developmental rate (Fig. 4.3A). Thermal breadth for developmental rate had no significant relationship with developmental time (r_6 = -0.66, p = 0.073), though there was considerable contraction in thermal breadth at intermediate events (appearance of eye spots and onset of cardiac function, Fig. 4.3A) by ~6°C and an overall reduction from early to late events by ~5°C (Fig. 4.3A). Similarly, thermal breadth for growth rate had no significant relationship with developmental time (r_6 = -0.62, p = 0.099), with thermal breadth greater in early events (gastrula to veliger) than late events (crawling to hatch) by ~3-4°C (Fig. 4.3B). Interestingly, thermal breadth was lowest for both traits at developmental events associated with the onset of cardiac function and the pigmentation of eye spots (Fig. 4.3).

4.5 | Discussion

Organismal development is a complex process, and its study requires a common currency in terms of the units used to assess its progress – this is typically time. However, temperature affects the speed of development, rendering time a problematic unit of measurement for comparative research. Here, measuring development in terms of the development of both form (growth) and function (developmental events) enabled me to i) produce thermal performance

curves specific to different periods of development enabling comparison of sensitivity throughout the embryonic period and ii) identify decoupling of the sensitivity of form and function during specific periods of development. Embryos of *L. stagnalis* were found to have progressively greater thermal sensitivity as they develop, evidenced by marked reductions in T_{opt}, CT_{max} and thermal breadth of thermal performance associated with development and growth.

Variation in thermal tolerance associated with advances in development has been observed in a number of taxa, including increases (Storch et al., 2011; Komoroske et al., 2014), decreases (Klockmann et al., 2016; Truebano et al., 2018; Ruthsatz et al., 2022) or with no statistical trend (Hall and Warner 2019) between stages. Reductions in thermal tolerance through development have been hypothesised to be driven by oxygen and capacity limited thermal tolerance (OCLTT) (Storch et al., 2011), but also the development of behavioural thermo-regulation at later stages, imposing a selection pressure on early immobile stages for enhanced thermal tolerance (Komoroske et al,. 2014), often termed the *Bogert* effect (Huey et al., 2003; Bowler & Terblanche, 2008). Conversely, increases in thermal tolerance between stages have been tied with body mass whereby larger, late stages could have reduced pressure from abiotic factors linked with high temperatures such as desiccation (Klockmann et al., 2016), but also differences in thermal strategy depending on the thermal environment that specific stages experience (Truebano et al., 2018; Ruthsatz et al., 2022). However, differences in methodology mean that studies are often assessing different aspects of thermal tolerance (Rezende et al., 2014), and thus tradeoffs between chronic vs acute tolerance are not fully elucidated (but see Truebano et

al., 2018). Increases in thermal tolerance through development are observed primarily in studies with acute assays (e.g. Klockmann et al., 2016; Ruthsatz et al., 2022), whilst reductions in thermal tolerance through development are observed in studies using chronic assays (e.g. Storch et al., 2011; Komoroske et al., 2014). Thus, such differences could be attributed to selection for enhanced chronic thermal tolerance early in development, but increased acute thermal tolerance late in development particularly if late developmental stages are able to behaviourally thermoregulate (Komoroske et al., 2014; Truebano et al,. 2018). Such a tradeoff could apply to the responses observed in the present study, though it is important to note that embryos were incubated to chronic temperatures throughout development and thus there is a confounding factor of thermal experience in these thermal sensitivity measurements. Therefore, increases in thermal sensitivity may also be driven by costs associated with chronic exposure, especially at extreme upper temperatures since heat damage is driven by both absolute temperature and duration of exposure (Rezende et al,. 2014).

Experiments with adult *L. stagnalis* have recorded rapid synthesis of HSPs in response to an acute thermal stress of 30°C (Foster et al., 2015). Consequently, it is possible that embryos under chronically elevated temperatures are undergoing overexpression of HSPs over the course of development and thus expending substantial resources to maintain continued embryonic development. Continued overexpression of HSPs has been tied with negative impacts on a number of traits, including growth, metabolism and fertility (Sorensen et al., 2003; Tomanek 2010) but also in influencing the expression of genes associated with other key physiological processes (Sorensen et al., 2005). Such overexpression of HSPs would form a

developmental time-dependent sensitivity to a chronic thermal challenge whereby increased costs culminate in reduced resources for later developmental stages (Kingsolver and Woods 2016), and thus drive reduced thermal tolerance in late embryonic stages.

Whilst T_{opt} and CT_{max} both reduced from early - late development in the present study, thermal breadth was lowest mid-development, at the commencement of heart rhythm and formation of eye spots. Such reduction in thermal performance at extreme, but sub-lethal temperatures could be attributed to the increased energetic costs associated with organogenesis at these periods in *L. stagnalis* embryonic development. Increased thermal sensitivity during organogenesis has been observed in a number of taxa, such as rainbow trout (*Oncorhynchus mykiss)* (Melendez and Mueller 2021) and lake whitefish (*Coregonus clupeaformis*) (Mueller et al., 2015). Such periods of heightened sensitivity are termed critical windows, consisting of a period in development where the phenotype of an organism has increased sensitivity to intrinsic or extrinsic factors (Burggren and Mueller 2015). Consequently, the periods of development associated with the start of heart rhythm and the formation of eye spots could represent critical windows in the development of *L. stagnalis*, and this is further reinforced by the sharp reductions in CT_{max} during these developmental periods observed in the present study. However, further experimentation would be required whereby thermal experience is removed as a conflating factor, to establish if these developmental periods are in fact critical windows.

Despite the widespread assessment of developmental responses using chronic thermal assays, few studies have investigated differences in responses throughout development in response to this commonly used assay. Here, significant differences were observed in thermal performance associated with each developmental event, for both growth and development in response to chronic incubation temperatures. Such empirical data have important implications not only for understanding developmental responses to temperature as a whole, but also in predicting developmental responses to changes in the natural environment. Many studies have used a single thermal performance curve derived from chronic assays to predict developmental responses to the fluctuating temperatures found in the environment but also to future climatic scenarios (e.g. Deutsch et al., 2008; Vasseur et al., 2014; Bernhardt et al., 2018; Khelifa et al., 2019); an approach often termed *rate summation* (Worner 1992; Kingsolver and Woods 2016). In this approach, there is an implicit assumption that responses to acute exposure are equivalent to those under chronic conditions (Niehaus et al., 2012), but also that thermal performance remains unchanged throughout development given the use of a single TPC (Kingsolver and Woods 2016). The limitations and inaccuracies of rate summation have been debated extensively (Niehaus et al., 2012; Sinclair et al., 2016; Kingsolver and Woods 2016), and the responses observed in the present study further call into question the use of this approach in predictive contexts. Further extensions to this approach that account for time-dependent changes in responses throughout an exposure period (e.g. Kingsolver and Woods 2016) could enable more accurate predictions of developmental responses to current and future climatic scenarios.

Here, I investigated whether thermal sensitivity differs during embryonic development in response to chronic incubation temperatures; a commonly used assay for characterising embryonic thermal performance. The use of automated, noninvasive imaging and downstream computer vision pipelines, enabled the extraction of longitudinal measures in both growth and development. Significant reductions in thermal performance were observed from early to late development, with earlier stages having significantly higher T_{opt}, CT_{max} and thermal breadth. Though the mechanisms underpinning these responses remain unclear, I highlight that these data are particularly relevant to the study and prediction of responses to fluctuating thermal environments and future climatic scenarios.

5 | Predicting developmental responses to fluctuating temperatures in embryos of *Lymnaea stagnalis*

5.1 | Abstract

Thermal performance curves (TPC) are amongst the most popular tools used to predict developmental responses to fluctuating temperatures and climate change. A popular approach involving the use of TPCs for development is rate summation, where a TPC constructed from responses to chronic exposure is integrated over time to predict cumulative thermal performance during a specific developmental period. However, key assumptions of rate summation are that chronic and acute exposure have the same physiological effects, and that thermal performance has no capacity to change, given that a single TPC is typically used to describe responses throughout development. Empirical studies of developmental responses to date have been limited in their ability to test these assumptions, primarily because developmental responses are typically measured as a cumulative effect at the end of a developmental process, rather than continuously throughout development. Here, I tested whether rate summation of TPCs based on responses to constant temperatures (15, 17.5, 20, 22.5, 25, 27.5, 30 and 32.5°C) could be used to accurately predict responses to fluctuating temperatures (17.7 ±7.7°C and 23.5 ±7.7°C) in embryos (n=120) of *Lymnaea stagnalis*. Embryos were imaged continuously throughout their development, with growth and developmental rate measured at key developmental events, enabling the investigation of how discrepancies between predicted and observed responses arise during development. I found that TPCs based

on constant temperatures could not accurately predict time to hatch or size at hatch in embryos developing in fluctuating temperature regimes. Comparison of responses during different developmental periods enabled the identification of altered developmental trajectories, with far earlier plateaus in growth at fluctuating as compared to constant temperatures. Furthermore, I found that embryos could grow over a far greater thermal range under fluctuating temperatures, indicating that chronic exposure does not have the same physiological effects as acute exposure. These results highlight the importance of considering responses throughout embryonic development, and caution the use of TPCs generated under constant temperatures to predict developmental responses to fluctuating temperature regimes.

5.2 | Introduction

Thermal variability across temporal scales, from seasons to days and even hours, is an inherent characteristic within most natural thermal environments and is thus experienced by even the most short-lived ectotherms (Kingsolver et al., 2015; Dillon et al., 2016). Yet, a large focus of thermal biology research has been on exposure to constant temperatures (CTs), with comparatively few studies on the effects of fluctuating temperature (FT) exposure (Niehaus et al., 2012). There have been increased efforts to resolve the responses of ectotherms to natural FT regimes (Vasseur et al., l2014; Sinclair et al., 2016), particularly because climate change continues to alter natural thermal environments with increases in the number and severity of extreme climatic events (IPCC 2023). However, predicting a species' vulnerability to climate

change is challenging given that both the thermal environment and thermal tolerance of an organism can vary depending on its ontogenetic stage (Kingsolver et al., 2011; Radchuk et al., 2013). Much of the theoretical literature has focused on the predicted responses of adult or mature life stages (Deutsch et al., 2008; Kingsolver et al., 2013; Vasseur et al., 2014), with early life stages often overlooked despite the considerable impact these stages can have on a species' vulnerability to climate change (Levy et al., 2019). For many species, early developmental stages constitute the most thermally sensitive life stage (Pandori and Sorte 2018; Dahlke et al,. 2020), and thus can cause life cycle bottlenecks that drive a species' vulnerability to climate change (Thomson et al., 2015; Levy et al., 2015). Early developmental stages are often immobile (e.g. encapsulated embryos) and have low thermal inertia due to their small size, making them more vulnerable to extreme temperatures than their adult counterparts (Woods et al., 2015; Huey et al., 1999). Moreover, maturation of regulatory functions for coping with temperature (Portner and Farrell 2008), pH (Conte 1984) and oxygen (Spicer 1995) occurs during development, and thus early developmental stages may be less able to cope with environmental challenge. As organisms develop, they accumulate energy reserves, potentially aiding in challenges associated with environmental stress, such as the cost of heat shock protein (HSP) synthesis (Pandori and Sorte 2018; Pottier et al., 2022). Consequently, there is a growing focus on resolving the vulnerability of early developmental stages to climate change (Pandori and Sort 2018; Levy et al., 2019; Pottier et al., 2022). A key component to this is understanding the responses of early developmental stages to the variable thermal environments many developing ectotherms experience.

A prevailing approach to predicting ectotherm responses to changes in environmental temperature involves the use of thermal performance curves (TPCs), which describe the relationship between temperature and rate of a given trait (Worner 1992; Angiletta 2009; Niehaus et al., 2012; Sinclair et al., 2016). When investigating developing organisms, TPCs are often quantified using chronic thermal assays to assess the effect of temperature on a given developmental outcome such as size, hatch time or survival (Angiletta 2009). Predicting responses to FT regimes is often achieved through rate summation, whereby the response at each temperature and time-step in an FT regime is described by the TPC, and the cumulative sum of these rates is used to quantify a given developmental outcome for the duration of exposure (Worner 1992; Niehaus et al., 2012). This approach minimizes the error arising from Jensen's inequality (Jensen 1906), which states that the function of the average (i.e. response at CTs) does not equal the average of the function (i.e. response to FTs) for functions with nonlinearity; an inherent property in ectotherm TPCs (Angiletta 2009). Rate summation is amongst the most popular approaches used in both empirical and theoretical studies for predicting biological responses to FTs (e.g. Worner 1992; Niehaus et al., 2012; Arrighi et al., 2013; Kingsolver et a., 2015; Bernhardt et al., 2018; Khelifa et al., 2019; Marshall et al., 2021) and future climate change scenarios (e.g. Deutsch et al., 2008; Vasseur et al., 2014). However, rate summation has not been without debate (Niehaus et al., 2012; Kingsolver and Woods 2016; Sinclair et al,. 2016; Levy et al., 2015).

Much of the debate surrounding the rate summation approach centres on several implicit assumptions, namely 1) that chronic and acute thermal exposure have the same physiological

effects (Niehaus et al., 2012), 2) the relationship between performance and temperature does not change throughout development (Kingsolver et al., 2015; Kingsolver and Woods 2016), and 3) that thermal experience does not alter the response to the current temperature (Kingsolver and Woods 2016). These implicit assumptions are not supported by the findings of many empirical studies in thermal biology (Niehaus et al., 2012; Sinclair et al., 2016), and thus have been proposed to drive the discrepancies between predicted and observed responses from empirical studies (Niehaus et al., 2012; Kingsolver et al., 2015; Sinclair et al., 2016). For example, it is widely reported that tolerance to extreme temperatures is inversely related with the duration of exposure (assumption 1), and thus non-lethal temperatures on short timescales can be lethal at longer timescales (Schulte et al., 2011; Rezende et al., 2014). Furthermore, thermal stress and acclimation can both change the shape and breadth of TPCs during development (assumption 2 and 3) (Niehuas et al., 2012; Kingsolver et al., 2015). Similarly, using a single TPC to describe responses throughout development goes against strong evidence of altered thermal sensitivity between developmental periods or stages (assumption 2) (Storch et al., 2011; Truebano et al., 2018; Ruthsatz et al., 2022). Though recent theoretical studies have attempted to address these limitations of rate summation (e.g. Kingsolver and Woods 2016; Roitberg and Mangel 2016), empirical data remains scarce. Specifically in developing ectotherms, the few studies that have assessed the efficacy of rate summation for predicting responses to FT exposure have found conflicting results. Accurate predictions of developmental time to FT exposure have been achieved via rate summation in embryos of the branchiopod, *Chydorus sphaericus* (Keen, 1978), early life stages of the armyworm, *Pseudaletia unipuncta* (Taylor and Shields 1990), Drosophila (Khelifa et al., 2019) and *Pieres napi* embryos (von

Schmalensee et al., 2021). However, a mismatch between predicted and observed responses has been found in growth and developmental rate for early life stages of the striped marsh frog, *Limnodynastes peronii* (Niehaus et al., 2012), growth rate in tadpoles of the toad, *Bombina orientalis* (Arrighi et al., 2013) and larvae of the tobacco hornworm, *Manduca sexta* (Kingsolver et al., 2015), and in estimating developmental rate in field conditions of fruit flies, *Drosophila melanogaster* (Khelifa et al., 2019). A key limitation of empirical studies that prevents further enquiry, is that despite development being a dynamic and continuous process, the traits that are studied are not measured continuously throughout development, but more often measured as a cumulative effect at the end of the developmental process (e.g. Niehaus et al., 2012; Arrighi et al., 2013; Kingsolver et al., 2015; Khelifa et al., 2019). This lack of temporal resolution prevents the identification of how discrepancies manifest during the course of development, and therefore testing alternative or additional hypotheses to rate summation are severely limited.

In this study, I tested whether rate summation of thermal performance curves derived from constant temperatures could accurately predict embryonic responses (growth and developmental rate) to FT regimes in the great pond snail, *Lymnaea stagnalis*. Embryos were incubated to repeating 24 h natural thermal regimes at two different average temperatures (17.7 ± 7.7°C and 23.5 ± 7.7°C), and imaged with an hourly resolution throughout their development (lasting 7 - 41 days post fertilization). Images were used to measure growth and developmental rate at 8 different periods of development, delineated by key development events (gastrula, trochophore, veliger, appearance of eye spots, first heartbeat, onset of

crawling, onset of radula function and hatch). Using developmental events as a scaffold permitted the investigation of how discrepancies between observed and predicted responses to FT regimes arise over the course of development, and thus enabled testing the following key questions: 1) can responses to acute exposure be accurately described by responses to chronic exposure and 2) does thermal experience of early developmental periods impact the response of later developmental periods?

5.3 | Materials and Methods

5.3.1 | Animal husbandry and embryo collection

Adult *Lymnaea stagnalis* (n = 25) were sampled from a laboratory stock population and transferred to individual, aerated aquaria (n = 25, vol. = 1.5 L, T = 20°C) filled with artificial pond water (APW) (chemical composition: $CaSO_4$ = 120 mg.L^{-1}, $MgSO_4$ = 245 mg.L^{-1}, $NaHCO_3$ = 192 mg.L^{-1}, KCl = 8 mg.L^{-1}) with a 12 h:12 h light:dark regime. The stock population at University of Plymouth was founded from a culture of the RENILYS strain (donated from IRNA, Rennes, France) established in October 2013 and maintained for multiple generations in large aquaria (n = 3, vol. = 25 L, T = 18-22°C, APW). Individuals of the culture used for the present study were fed lettuce *ad libitum* and water changes were carried out 1-2 times per week. Egg masses were removed from aquaria using double layer laminate plastic within 24 h of deposition, and inspected under a dissecting microscope (10 – 40 ×, HM-4, Microtech); egg masses containing embryos that had not developed beyond the 2nd cell division were extricated from the egg mass

and evenly distributed between two 96-well microtitre plates (Nunc, Microwell, 96 wells, 350 µl well^{-1}) held at one of the two FT regimes ('Cold': 17.5 ± 7.7°C, 'Warm': 23.5 ± 7.7°C).

5.3.2 | Thermal assays and bio-imaging

Lymnaea stagnalis embryos (n=120) were incubated under repeating 24h thermal regimes at two mean temperatures: 17.5°C (n=60) and 23.5 (n=60). The pattern of these simulated FT regimes was replicated at two mean temperatures (17.5 ± 7.7 or 23.5 ± 7.7 °C) from a 24h thermal log of a freshwater stream on Dartmoor (50.465493, -4.036087)(Fig. 5.1). Embryos were imaged for 20s (1024x1026 pixels, 30 frames sec^{-1}, 8-bit depth, 200x magnification) from 1st cell division until hatching. Imaging was performed using OpenVIM video-microscopes with robotic XY motorized stages for high-throughput time-lapse imaging of developing embryos (Tills et al., 2018), housed in a controlled temperature room (T = 16°C) . Microtiter plates (Nunc, Microwell, 96 wells, 350 µl well^{-1}) containing embryos were held within incubation chambers (H101-K-Frame, Okolab, Naples, Italy) of two separate OpenVIM systems, each corresponding with a different thermal regime (17.5 ± 7.7 or 23.5 ± 7.7 °C). OpenVIM instruments were controlled using MicroManager, an open source software (Edelstein et al 2010) for performing acquisition of videos of individual embryos via a sequence of individual TIFF files. A Python script was used to concatenate TIFF files into single timelapse videos for individual embryos to facilitate subsequent analysis (see https://www.github.com/EmbryoPhenomics/fast_img_compile/).

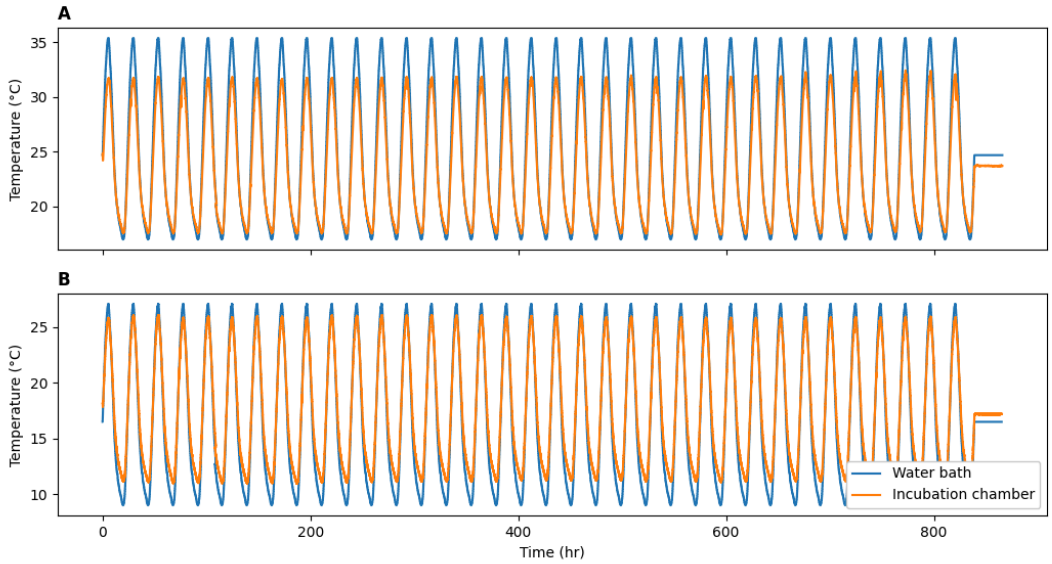

Figure 5.1. Simulated repeating 24 hr natural thermal regimes, at two mean temperatures: 23.5°C (A) and 17.5°C (B), both with an amplitude of ± 7.7 °C. Water bath temperatures are included to show the offset required to achieve the desired incubation chamber temperatures for a room ambient of 16°C.

Temperature within the incubation chamber was controlled via water circulation from a water bath (H101-CRYO-BL, Okolab) with humidified air supplied by an air pump (OKO AP, Okolab) and a Dreschel bottle situated within the water bath. Dynamic control of the water bath was achieved using a programmable temperature control unit (Bold Line Temperature Unit, OkoLab) in conjunction with DATA LOG software (Okolab) running on a PC running Windows 10. Dynamic temperature control was achieved using the 'chamber' mode, i.e. directly setting the water bath temperature, rather than the 'sample' mode, i.e. a feedback loop is used to adjust the water bath temperature to achieve the setpoint temperature in the incubation chamber. Bypassing the feedback loop enabled more precise and rapid temperature change, though it requires accounting for the offset between the temperature in the water bath and the incubation chamber. To identify the offsets required for a room ambient of 16°C, I initially used the 'sample'

mode to identify the offsets at the minimum and maximum temperature of both FT regimes. These offsets were then used to linearly re-scale the thermal regimes to new minimum and maximum temperatures and thus account for appropriate offsets throughout each regime (Fig. 5.1). Finally, water levels within microtitre plates were checked daily and topped up with distilled water as required.

5.3.3 | Image analysis

Individual models trained using deep learning were used for measuring (i) egg location in the image, (ii) size of embryos (area), and for (iii) ascertaining the timing of specific developmental events. Training and inference of each of the deep learning models was performed using an RTX 3090 GPU (NVIDIA, Santa Clara, USA).

i) Egg location was determined using an Xception model (Chollet 2017), pretrained using a manually annotated dataset generated in a prior study (see Chapter 4), to localise individual eggs in each video. The location of the bounding box containing the embryo was used to filter video footage for all subsequent analytical steps (see https://github.com/EmbryoPhenomics/egg_detection).

ii) Embryo size was measured using a pre-trained *UNet* (Ronneberger et al., 2015) model to perform semantic segmentation, enabling embryo area to be measured via pixel counts (see https://github.com/EmbryoPhenomics/embryo_segmentation/), with subsequent conversion to mm^2 using a separate function (outlined in Chapter 4.3.3).

iii) Overall developmental time (1st cell division to hatch) was manually quantified with an hourly resolution and then key developmental events were identified through the use of *Dev-ResNet*: a 3D convolutional neural network for automated detection of developmental events (see Chapter 3 - Ibbini et al., 2024).

5.3.4 | Prediction of developmental responses

Prediction of embryonic responses to FTs were made using the implementation of rate summation (Tanigoshi and Logan 1979; Worner 1992; Niehaus et al., 2012) which uses thermal performance curves derived from responses to constant incubation temperatures. As described in Chapter 4, *L. stagnalis* embryos were imaged (1024x1026 pixels, 30 frames sec-1, 8-bit depth, 200x magnification) every hour from 1st cell division until hatching at a range of constant temperatures (15°C (n = 58), 17.7°C (n = 60), 20°C (n = 38), 22.5°C (n = 55), 25°C (n = 38), 27.7°C (n = 59), 30°C (n = 37) and 32.5°C (n = 60). This was achieved using the OpenVIM system (Tills et al., 2018), where embryos were held in 96-well microtiter plates housed within jacket incubation chambers maintained at one of the given constant temperature treatments.

The two key traits of interest for the present study were developmental rate (h^{-1}) and growth rate ($mm^2\ h^{-1}$), both measured (1) throughout the entire embryonic period and (2) in the periods between 1st cell division and particular developmental events. Thermal performance curves for both growth and developmental rate for periods (1) and (2) were then quantified using the rTPC (Padfield et al., 2021) package in R v4.3.1, where four different TPC functions

(Gaussian, Weibull, Thomas et al., 2017 and Rezende et al., 2019). Two of these functions – the Gaussian and Weibull functions – have been commonly used to theoretically or empirically describe thermal reaction norms (Huey and Kingsolver, 1993; Palaima and Spitze, 2004), whilst the latter functions of Thomas et al., (2017) and Rezende et al., (2019) were selected because their complex structure should provide a better fit to nonlinear data. The model proposed by Rezende et al., (2019) provided the best fit across the most developmental events for both traits (Table A2 and A3), developmental and growth rate, and in cases where this was not the case, the difference in AIC between the best ranking model was minimal. Thus, this model was used for comparisons of thermal performance metrics between developmental events.

Embryonic responses were predicted using rate summation (Tanigoshi and Logan 1979; Worner 1992) which is described as:

$$\sum_{t}^{N} r(T(t)),$$

where *r* equals the incremental rate of growth or development as described by the corresponding TPC (Fig. 5.2A, G), and *T* equals the mean temperature at time interval *t*. For predictions of the timing of developmental events, rate summation was performed until the cumulative sum of incremental rates equals one (Fig. 5.2D). Conversely, for size predictions, rate summation was performed until the predicted time of a developmental event (Fig. 5.2G-I). Rate summation only predicts the mean response to a FT regime, given that one typically uses a TPC fitted to mean thermal performance observed at constant temperatures (Worner 1992). To enable statistical comparisons to be made between predicted and observed responses to FTs, it

was necessary to generate artificial individual-level variation around the predicted responses for different temperatures (Lerin 2004). This was achieved using a Weibull cumulative distribution function (CDF), a commonly used method to describe inter-individual variation in invertebrate developmental time (Wagner et al., 1984). The Weibull function is used because cumulative distributions of invertebrate developmental time are typically asymmetrical around the median, often skewed to the right (Wagner et al., 1984). The Weibull CDF is not fitted to absolute developmental time, but instead to 'physiological time', where individual-level times are divided by the treatment-specific median developmental time (see Wagner et al., 1984). The use of physiological time enables standardised comparisons across different treatments, with differences in the overall rates of development (Wagner et al., 1984; Fig. 5.2E). Weibull CDFs have previously been performed with terrestrial insect data (Wagner et al., 1984; Lerin 2004), though the same approach was effective ($r^2 = 0.96$) when applied to hatch time data of *L. stagnalis* embryos (Fig. 5.2E) and to developmental event timings (Fig. A7). Once the Weibull CDF is applied, inter-individual variation is generated by randomly sampling the CDF and multiplying samples by the predicted mean developmental time for a given developmental period, to get absolute developmental times (Lerin 2004; Fig. 5.2F). Variation in size estimation was introduced by performing rate summation for the duration of each individual sample time predicted in the preceding step (Fig. 5.2I).

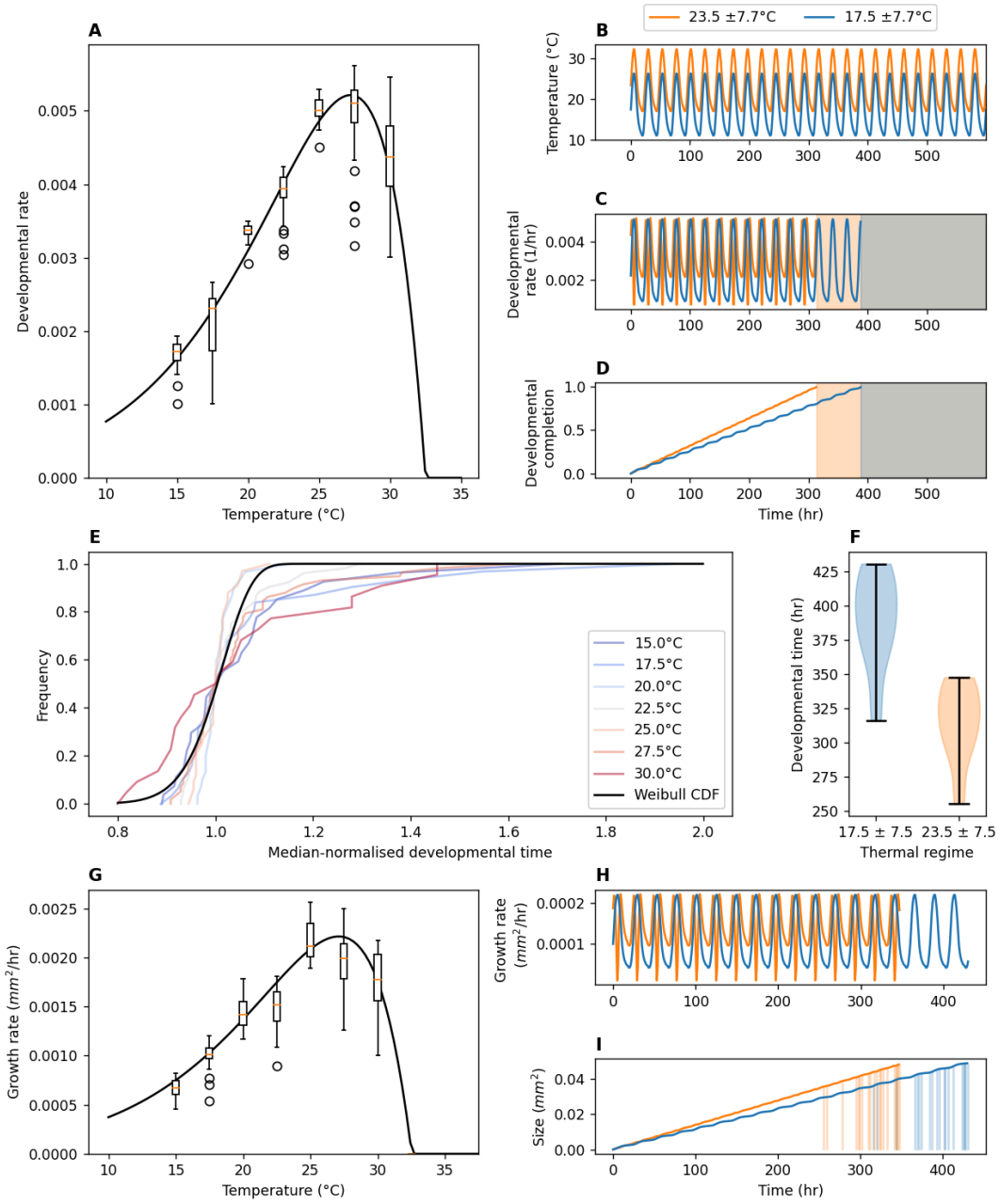

Figure 5.2. A TPC fitted to average developmental rate (A, solid black line), constructed from responses of *Lymnaea stagnalis* embryos incubated to constant temperatures (15°C (n=58), 17.5°C (n=60), 20°C (n=38), 22.5°C (n=55), 25°C (n=38), 27.5°C (n=59), 30°C (n=37) and 32.5°C (n=60)) (A), is used to predict hatch time using rate summation for the two fluctuating temperature regimes 17.5 ± 7.7 and 23.5 ± 7.7 °C used in the present study (B). The point at which a cumulative sum of incremental developmental rates (C) equals one (D) is determined to be the predicted

time to hatch (C-D). The start of shaded regions in (C) and (D) correspond to the time at which embryos are predicted to have hatched, with colouration corresponding to the two fluctuating temperature treatments outlined in (B). Median-normalised (i.e. physiological time, Wagner et al., 1984) cumulative distribution curves for hatch times observed at constant incubation temperatures were used to fit a Weibull CDF (E) for the purpose of introducing variation (samples = 20) around hatch time predicted using rate summation (F). Finally, a TPC fitted to average growth rate (G, solid black line), constructed from responses of *L. stagnalis* embryos incubated to constant temperatures was used to predict mean size at hatch via rate summation, where a cumulative sum (I) of incremental growth rates (H) was performed for the duration of developmental time predicted in (C-D). Variation in size at hatch is introduced by performing rate summation for the different developmental times predicted in (F). The vertical lines in (I) correspond to the different hatch times predicted in (E-F), and the intersection with the size curves illustrate the size variation introduced.

5.4 | Results

5.4.1 | Rate summation with a single TPC does not accurately predict size and time at hatch

A single TPC approach to rate summation was used to predict time to- and size at- hatch for two different FT regimes. This approach is commonly used throughout the literature for predicting developmental responses to FT environments and future climatic change scenarios (Tanigoshi and Logan 1979; Worner 1992; Deutsch et al., 2008; Niehaus et al., 2012; Vasseur et al., 2014; Sinclair et al., 2016). Embryos incubated to both 17.5 ± 7.7 °C and 23.5 ± 7.7 °C FT regimes hatched significantly later (17.5 ± 7.7 °C: t_{21} = -9.114, $p < 0.0001$, 23.5 ± 7.7 °C: t_{21} = -9.137, $p < 0.0001$, Fig. 5.2A) and at a smaller size (17.5 ± 7.7 °C: t_{21} = -8.714, $p < 0.0001$, 23.5 ± 7.7 °C: t_{21} = -6.446, $p < 0.0001$, Fig. 5.3B) than predicted on the basis of rate summation alone, indicating

that a single TPC derived from constant temperature exposure overestimates growth and developmental rate in dynamic FT regimes for embryonic development in *L. stagnalis*.

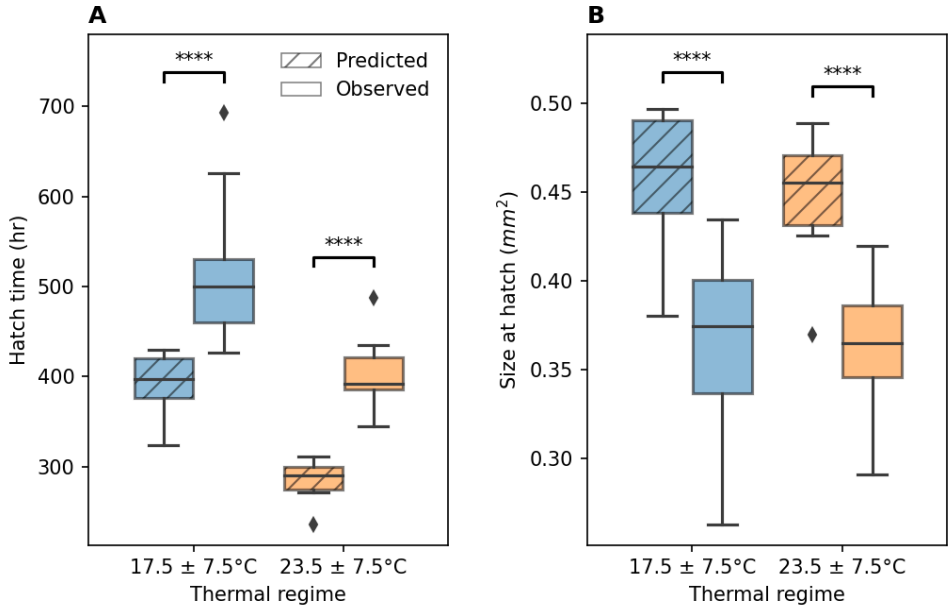

Figure 5.3. Comparison between predicted responses, quantified from rate summation of TPCs derived from constant temperature incubation, and observed responses for development (A) and growth (B) in *Lymnaea stagnalis* embryos incubated to two fluctuating temperature regimes: 17.7 ± 7.7 (n = 22) and 23.5 ± 7.7 °C (n = 11). The sample size is the same for artificial samples in predicted responses. Statistical significance between predicted and observed data was determined via paired t-test (****: $p < 0.0001$).

5.4.2 | Discrepancies between predicted and observed embryonic responses are primarily driven by late developmental events

A developmental event approach was used to perform standardised comparisons between predicted and observed responses during different periods of embryonic development in *L. stagnalis* (Fig. 5.4). This approach enabled identification of the differential contribution of

periods of development to overall discordance between observed and predicted time and size at hatch.

Though both growth and development rate were accurately predicted for some early to intermediate developmental events (Fig. 5.4, Table A4), there were considerable discrepancies in late developmental periods (crawling - hatch, events 6 - 8) in terms of both time of- and size at- onset of developmental events associated with crawling, radula function and hatching (Fig. 5.4B-E, crawling - hatch, events 6-8, Table A4). The magnitude of discrepancies between predicted and observed responses changed non-linearly as development progressed (Fig. 5.4B-E)

Lymnaea stagnalis embryos exhibit considerable growth throughout development, with growth typically not plateauing until later developmental events (typically after the onset of radula function) when the embryo occupies >80 % of the egg (Fig. 5.5). Consequently, size curves can be informative for comparison of developmental trajectories between temperature regimes. Here, size curves were produced using hourly measurement of embryo area. Comparing size curves between the two FT regimes revealed marked differences in growth trajectories, with growth slowing considerably after the onset of cardiac function in the 23.5 ± 7.7 °C regime, whilst not slowing until substantially later in the 17.5 ± 7.7 °C regime, just prior to the onset of radula function (Fig. 5.5). The growth trajectories in both FT regimes indicate far earlier plateaus in growth than in constant temperature incubation in which growth only slows considerably after the onset of radula function, just prior to hatching (Fig. 5.5).

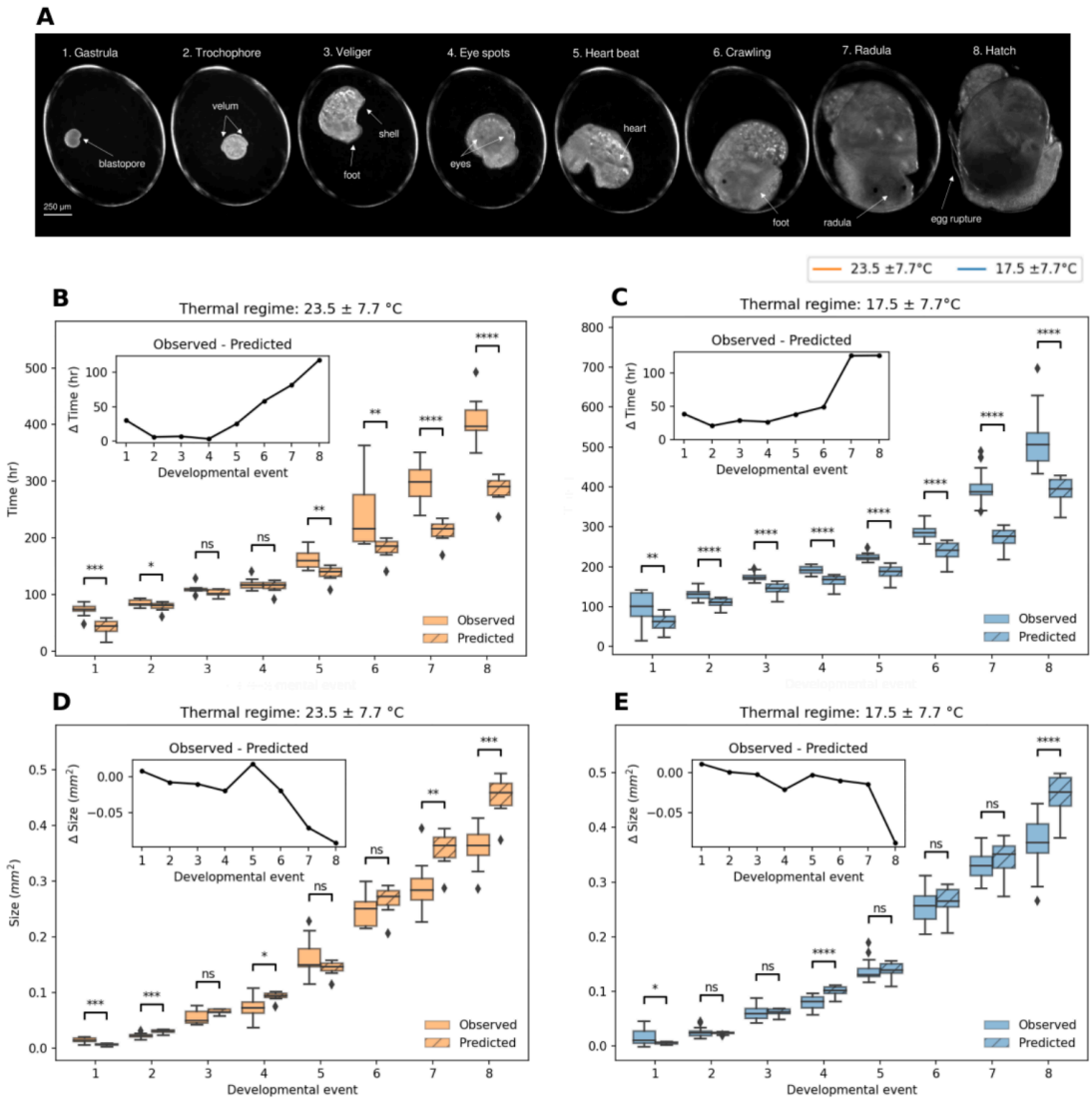

Figure 5.4. Comparison between predicted (quantified from rate summation of TPCs derived from constant temperature incubation) and observed developmental responses for development (B-C) and growth (D-E) at key developmental events during *Lymnaea stagnalis* embryonic development (A). Embryos were incubated to two different fluctuating thermal regimes: 17.7 ± 7.7 (n = 2 2) and 23.5 ± 7.7 °C (n = 11). The sample size of predicted samples generated using a Weibull cumulative distribution function (Fig. 5.2E-F) is the same as those in empirical data. Event-specific predictions are quantified via performing rate summation from 1st cell division to a given event, and are not the product of performing rate summation between consecutive events. Statistical significance

between predicted and observed data was determined via paired t-test (****: $p < 0.0001$, ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, ns: $p > 0.05$). Absolute deltas between mean predicted and observed values are shown in the subplots (B-E).

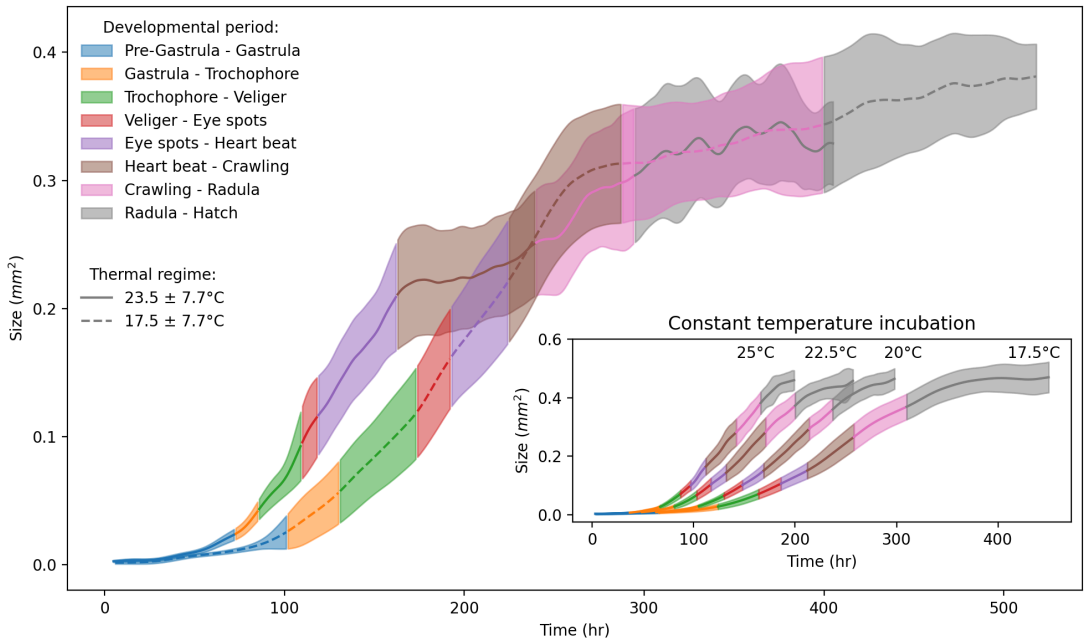

Figure 5.5. Size curves (mean ± sd) with developmental periods (i.e. period between consecutive developmental events) overlayed for both fluctuating temperature treatments used in the present study (17.7 ± 7.7 (n = 22) and 23.5 ± 7.7 °C (n = 11)). Mean size curves derived from constant temperature treatments with comparable average temperatures (17.7°C (n = 60), 20°C (n = 38), 22.5°C (n = 55), 25°C (n = 38)) to the fluctuating temperature regimes of the present study are displayed in the subplot.

5.4.3 | Developmental responses to acute exposure are not accurately described by responses to chronic exposure

Rate summation most commonly involves the use of TPCs derived from constant temperatures (e.g. Fig. 5.2A, G), and therefore a key implicit assumption in this approach is that chronic

exposure has the same physiological effects as acute exposure. To test the validity of this assumption, I quantified hourly growth rates during the exponential growth phase (Fig. 5.6A, C) for embryos incubated to both FT regimes, as this is the period of growth where size is most uniform regardless of embryo orientation, facilitating measurement of hourly growth rate. Increasingly asymmetric morphology in later stages leads to substantial variation in area measurements depending on embryo orientation (Fig. 5.6A, C), and thus measuring hourly changes in size becomes limited. The relative length of this exponential growth phase in *L. stagnalis* typically starts from gastrulation and ends at the onset of radula function at chronic incubation temperatures (Fig. 5.5). This exponential growth phase is shortened to just prior to the onset of crawling at 17.5 ± 7.7 °C (Fig. 5.5, Fig. 5.6A), but ends considerably earlier at 23.5 ± 7.7 °C regime, just after the onset of cardiac function (Fig. 5.5, Fig. 5.6C). Consequently, the relative length of this exponential phase used for quantification of hourly growth is different between the two regimes and this is highlighted by the shaded gray region in Fig. 5.6A, C.

Hourly growth observed under both FT regimes, was compared with predicted hourly growth as described by the TPC for overall embryonic growth in *L. stagnalis* (Fig. 5.2G-I). There were significant discrepancies between observed and predicted hourly growth rates in both 17.5 ± 7.7 °C and 23.5 ± 7.7 °C regimes (Tukey's HSD, $p < 0.05$, Fig. 5.6B, D). Moreover, growth at upper extreme temperatures in the 23.5 ± 7.7 °C regime was significantly greater than would be predicted on the basis of rate summation alone (Fig. 5.6D), indicating a much greater temperature range over which growth can occur in response to acute thermal exposure than to chronic thermal exposure in embryos of *L. stagnalis*.

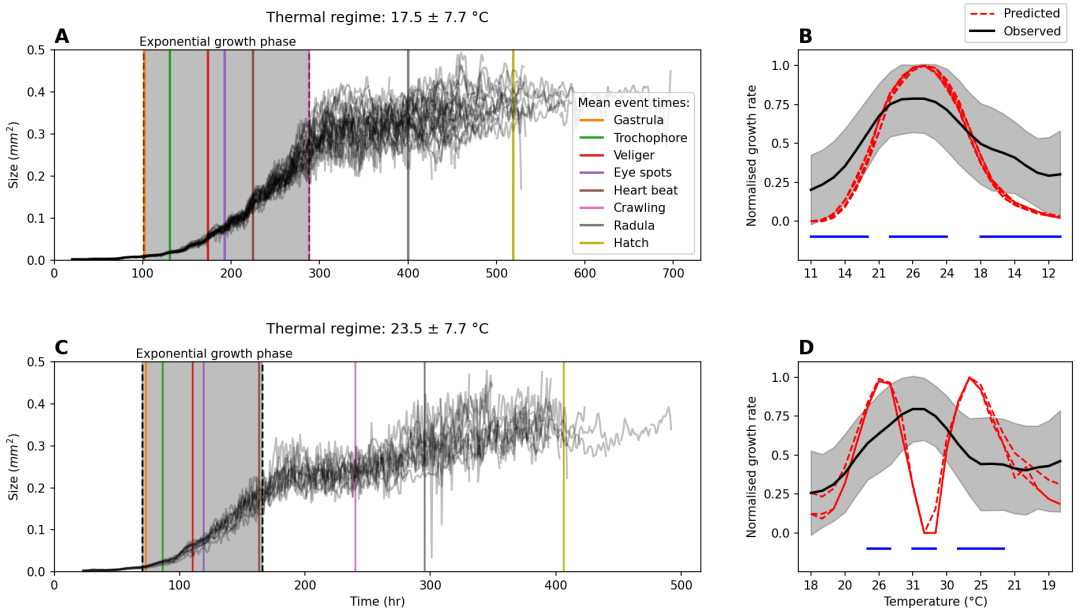

Figure 5.6. Individual size curves (A, C: gray lines) throughout embryonic development of *L. stagnalis* exposed to two fluctuating temperature regimes: 17.7 ± 7.7 (A-B) (n = 22) and 23.5 ± 7.7 °C (C-D) (n = 11). The exponential growth phase (A, C, shaded grey region between dash black lines) for embryos in both treatments is used for measuring hourly growth rate (mean ± sd) in response to acute temperature change (B, D), with comparison with predicted growth rates (C, D) based on rate summation of a TPC constructed from chronic temperature exposure (Fig. 5.2G). Both observed and predicted hourly growth rates are normalised for statistical comparison. Horizontal blue lines in (B,D) correspond to temperatures where there are significant differences between observed and predicted growth rates as determined by repeated measures ANOVA (Tukey's HSD, $p < 0.05$).

5.5 | Discussion

Researchers have long attempted to use responses to constant temperatures to predict and understand responses to FT environments, most often using rate summation of thermal performance curves (Worner 1992; Lerin 2004; Niehaus et al., 2012; Kingsolver and Woods

2016; Sinclair et al,. 2016). There is an increasing number of empirical studies that have tested these predictive approaches, though with conflicting results in terms of the efficacy of rate summation (Niehaus et al., 2012; Kingsolver et al., 2015; Khelifa et al., 2019; Marshall et al., 2021). A key limiting factor in such empirical studies is the use of single endpoints with no measures for how responses change during an exposure period, therefore limiting understanding of how discrepancies between predicted and observed responses emerge during development. Consequently, here the efficacy of the rate summation approach was tested in predicting developmental responses to FTs, using TPCs derived from exposure to CTs, leveraging high-throughput imaging throughout development to identify where discrepancies arise in predictions over the course of embryonic development in the great pond snail, *Lymnaea stagnalis.* Embryos developed significantly slower and hatched significantly smaller than would be predicted on the basis of rate summation of single, whole-developmental TPCs. Unpacking these responses to key developmental events throughout embryonic development enabled the identification of developmental plasticity from differences between embryos developing under FT and CT environments. Embryos plateaued in growth earlier in the developmental itinerary under FTs and took longer to fully occupy the egg capsule and hatch. Altered developmental trajectories could be a driver behind the discrepancies observed in other empirical studies testing rate summation, and the acquisition of longitudinal data in such studies would help determine where fluctuating temperatures elicit developmental responses beyond those described by passive plasticity. Below, I discuss some possible explanations for the patterns observed in the present study, and consider how they compare with the wider empirical literature surrounding fluctuating temperatures.

Growth trajectories in embryonic development of *L. stagnalis* in controlled laboratory conditions follow an exponential curve with a relatively short plateau in growth towards the end of development as the embryo fully occupies the egg capsule prior to hatching (e.g. Bandow and Weltje 2012; Liu et al., 2022), similar to the CT responses observed in the present study. Considering that embryos exposed to the FTs used in the present study hatch at a smaller size than those incubated to CTs, this could indicate that the earlier plateaus in growth observed in FT regimes are a mal-adaptive plastic response (Morris and Rogers 2013), given that increased size is typically linked with greater fitness among a population or species (Kingsolver and Huey 2008). This is further supported by the fact that growth did not proceed any more rapidly at early developmental periods in FT regimes compared to CT regimes. Acceleration of growth and development might be expected under FT regimes given that increased fluctuations in temperature might signal deteriorating conditions with risks such as desiccation (Semlitsch and Wilbur 1988; Niehaus et al., 2012). However, note that embryos used in the present study were sampled from a population maintained in the laboratory across many generations, and thus, tolerance to FT conditions could be substantially reduced over a field population. For example, in an empirical study with larvae of the tobacco hornworm, individuals from a laboratory population were found to have reduced tolerance and higher sensitivity to warmer FTs than a field population (Kingsolver et al., 2020). Further study with different field populations of *Lymnaea stagnalis* is required to confirm whether the responses found in the present study to FTs are specific to this laboratory population, or are indeed representative of species responses

as a whole. The methods developed and applied here, should prove an excellent foundation for such studies.

There are conflicting results in the empirical literature surrounding the efficacy of rate summation based on CT responses. It has proven an effective approach in predicting population growth rates in green alga, *Tetraselmis tetrahele* (Bernhardt et al., 2018), developmental rates for a range of *Drosophila* species (Khelifa et al., 2019) and also in the butterfly, *Pieries napi* (von Schmalensee et al., 2021). Though it has been ineffective in predicting growth in tadpoles of the toad, *Bombina orientalis* (Arrighi et al., 2013), larval growth in the tobacco hornworm, *Manduca sexta* (Kingsolver et al., 2015), growth and development in tadpoles of striped marsh frogs, *Limnodynastes peronii* (Niehaus et al., 2012), and growth, survival, feeding rates and byssal thread production in bay mussels, *Mytulis trossulus* (Marshall et al., 2021). A key limitation proposed by these latter studies is the implicit assumption for performance under chronic exposure to be equivalent to acute exposure (Niehaus et al., 2012; Kingsolver et al., 2015), which can lead to either over- or under-estimation of responses to FTs as a result of either acclimation or thermal stress respectively. Moreover, exposure to extreme temperatures in fluctuating conditions may have lasting effects and thus instantaneous responses as predicted by rate summation would fail to capture such effects (Arrighi et al., 2013). The empirical data of the present study support both of these hypotheses: performance under acute exposure was found to not be equivalent to performance under chronic exposure, evidenced by the under-estimation of growth rates to extreme upper temperatures in the 23.5 ± 7.7°C treatment. Moreover, there were clear lasting effects as a consequence of FT incubation as evidenced by

the significantly different growth trajectories in late developmental windows in both FT regimes in comparison to growth trajectories to constant incubation temperatures. These results - together with those of Chapter 4, where thermal performance under chronic exposure changes over the course of development - call into question the use of responses to constant temperatures to describe those to fluctuating temperatures. Clearly a more dynamic predictive approach is required, where there is capacity for the patterns in developmental responses observed here and elsewhere, to be integrated into final predictions. However, an important precursor to the development of such an approach is a more thorough characterisation of developmental responses under different fluctuating thermal regimes, outside of the small number of regimes used in empirical studies to date, the present study included.

Thermal performance curves, and by extension rate summation, remain important tools to understand and predict developmental responses to current and future variable thermal environments. Here, using a developmental event approach it was shown that rate summation of CT responses can result in inaccurate predictions of responses to FTs, primarily as a consequence of altered developmental trajectories at FTs. Altered developmental trajectories could explain the often marked differences between predicted and observed responses found in other empirical studies using FTs. Nevertheless, rate summation could still be a useful approach for modelling developmental responses for each stage in embryonic development, evidenced by the accurate predictions for some developmental events found in the present study. However, a more comprehensive approach that encapsulates the trajectories that embryonic development may follow in variable conditions would be required for better predictions of responses during

embryonic development as a whole. A more thorough characterisation of developmental responses and trajectories to FT regimes that differ in their characteristics (e.g. average temperature and amplitude of fluctuations) would allow a more complete understanding of how ectotherms develop in variable thermal environments, but also potentially to future climatic scenarios.

6 | Conclusion

6.1 | Thesis recap and introduction

The main aim of my thesis was to measure embryonic responses under constant temperatures in a TPC framework, and to test the capacity of these measurements for predicting responses to fluctuating temperature regimes using rate summation. Rate summation remains perhaps the most commonly used predictive approach in the fluctuating temperature (FT) literature (see Sinclair et al., 2016), particularly for predicting biological responses to climate change scenarios (e.g. Deutsch et al., 2008; Vasseur et al., 2014; Levy et al., 2019). However, there has been increased debate regarding the efficacy of the approach, particularly because there is a growing number of empirical studies that find rate summation ineffective at predicting developmental responses to FT regimes (e.g. Niehaus et al., 2012; Arrighi et al., 2013; Kingsolver et al., 2015; Marshall et al., 2021). A key limitation of prior empirical studies was their focus on developmental endpoints (e.g. size at hatch)(Worner 1992; Niehuas et al., 2012; Kingsolver et al., 2015; Khelifa et al., 2019), with no measurement of the responses that lead to those endpoints (e.g. growth during development). Consequently, this thesis focused on developing automated tools for the characterisation of responses throughout embryonic development, namely the quantification of cardiac rhythm (Chapter 2 - Ibbini et al., 2022), detection of developmental event timings (Chapter 3 - Ibbini et al., 2024) and localisation of eggs and measurement of size (Chapter 4). Comparing embryonic responses between developmental periods enabled me to identify several key assumptions of the rate summation approach that I

have shown to be invalid: 1) the relationship between performance and temperature does not change over the duration of exposure or throughout development (Chapter 4), 2) chronic and acute thermal exposure have the same physiological effects (Chapter 5), and 3) that thermal experience does not alter the response to the current temperature (Chapter 5).

In this final chapter, I aim to synthesise the main findings of my thesis, discuss their implications and outline avenues for future work. The below discussion is structured into two sections: in the first (6.2), I evaluate the methodological approaches used in my thesis and their contribution to the findings of my thesis, and in the second (6.3), I discuss the contribution of my thesis to our knowledge of developmental responses to fluctuating temperatures.

6.2 | Computer vision enabled approaches to assessing developmental responses to fluctuating temperatures

Embryonic development is a process occuring over a range of temporal scales, and therefore a key limiting factor in its study can be the sampling strategy. Approaches typically rely on sampling at fixed time-points, with significant associated limitations on understanding the continuous process of development. Such an approach is particularly limiting in considering the impacts of fluctuating temperatures on organismal development, given the inherent temporal component of both the predictor and response variables. Consequently, a key prerequisite for the research in my thesis was the ability to produce longitudinal data, i.e. high-resolution time series data, for individual embryos alongside different temperature regimes.

In my thesis, I used a high-throughput imaging platform (OpenVIM, Tills et al., 2018) to record the development of aquatic embryos from 1st cell division to hatch. Manually extracting longitudinal phenotypic data from such video would have represented a significant bottleneck (Zhou and Wong 2006; Cardona and Tomancak 2012; Meijering et al., 2016), and this could explain why discontinuous non-longitudinal approaches have been standard to date in the empirical literature for fluctuating temperatures. Therefore, the initial phase of my PhD was focused on developing automated approaches to bioimage analysis for the extraction of longitudinal data from the large video datasets produced. Below, I will discuss the rationale behind using deep learning methods for automating bioimage analysis, as opposed to traditional computer vision techniques.

With bioimage datasets becoming increasingly easy to acquire, as the barriers to the use and adoption of imaging hardware reduce both in terms of cost and availability (Baden et al., 2015; Chagas et al., 2017; Tills et al., 2018), the field of bioimage analysis has seen rapid growth to create methods for effectively analysing these datasets (Swedlow et al., 2009; Walter et al., 2010; Meijering et al., 2016). The challenge of bioimage analysis is that there is considerable variation in bioimage datasets, due to the myriad different imaging modalities, sample preparation protocols, acquisition parameters and species or life stage of study (Hallou et al., 2021). Traditional computer vision methods, involving hand-crafting image filters and identifying effective combinations of lower level algorithms such as image thresholding, have been effective in specific domains but have often been limited in terms of transferability

(O'Mahony et al., 2019), given that images between species, life stages or even experimental setups can have substantial differences. In Chapter 2, I present a transferable computer vision tool (HeartCV, Ibbini et al., 2022) for automated localisation and quantification of cardiac rhythm in videos of transparent animals, and it was effective across a range of taxa with marked differences in their cardiac morphology. Though there are other transferable computer vision methods based on traditional algorithms, such as EthoVision for video tracking of animals (Noldus et al., 2001), there is a scarcity of transferrable methods for the majority of observable phenotypic traits due to the considerable effort required to develop a computer vision pipeline that is effective across applications. Moreover, for many traits that differ considerably in their observable characteristics between species and/or life stages, such as size, a transferable approach that relies on traditional computer vision algorithms is likely to be intractable. Deep learning - where neural networks composed of multiple layers of artificial neurons are trained to learn representations of data that produce the most accurate predictions - has gained significant traction in its ability to overcome this issue, given that image filters are automatically learnt during the training process, and are not manually developed (LeCun et al., 2015; O'Mahony et al., 2019). The only prerequisites to training a deep learning model is sufficient training data and computing hardware, requirements that are increasingly easy to meet given the prevalence of manual annotation tools (Dutta and Zisserman 2020; Hollandi et al., 2020) and continued improvement in hardware performance (LeCun et al., 2015). This has resulted in widespread adoption of deep learning approaches in bioimage analysis (Hallou et al., 2021; Nogare et al., 2023), and has facilitated novel avenues for analysing bioimaging data beyond that of traditional computer vision, such as visualising and characterising the continuous

process of embryonic development (e.g. Toulany et al., 2023; Ibbini et al., 2024). Below, I will unpack the contributions of the deep learning components of my thesis to the automation of bioimage analysis, both within the context of my thesis and the wider literature of experimental biology.

Automated approaches for bioimage analysis using deep learning

Two key traits that have been commonly used in the empirical and theoretical literature for fluctuating temperatures include growth and developmental rate (Niehaus et al., 2012; Kingsolver et al., 2015; Sinclair et al., 2016). To extract these traits in a longitudinal fashion from the video recorded using OpenVIM (Tills et al., 2018), I developed and trained deep learning models for automated localisation of *L. stagnalis* eggs (Chapter 4), embryonic size estimation (Chapter 4) and detection of developmental events (Chapter 3 - Ibbini et al., 2024). Below I will discuss each of these further.

A common barrier in the application of DL models to non-model species is the lack of public training datasets, and thus the creation of training data constitutes a considerable challenge given that the length of time an expert(s) can allocate to making ground-truth annotations is limited. Consequently, small training datasets (100s rather than millions of samples) are not uncommon in the biological sciences, and was the case for the datasets created in this thesis for both localisation of eggs and size estimation of embryos. Therefore, a key focus in this thesis was in training accurate DL models on small training datasets. Here, I successfully trained 2D convolutional neural networks (CNNs) on very small imaging datasets (~ 300 ground truth

images) for both localisation of eggs and size estimation of embryos, demonstrating the robustness of 2D CNNs in producing accurate predictions with limited training data. Furthermore, the 2D-CNN models used here were originally developed for automating computer vision tasks that are completely different to that of my thesis, such as cell and nuclei segmentation (U-Net, Ronneberger et al., 2015) and classification of common objects in general digital image datasets such as ImageNet and CIFAR-10 (He et al., 2015). The application of such models to the localisation and segmentation tasks in my thesis highlight the versatility of 2D-CNNs in automating some of the most common image-related tasks posed to developmental biologists with minimal human input.

For the detection of developmental events (Chapter 3 - Ibbini et al., 2024), I developed a small and efficient 3D convolutional neural network, Dev-ResNet, consisting of just ~5.2 million parameters and ~3.7 GFLOPs respectively (Giga floating point operations), capable of integrating both spatial and temporal information to detect developmental events. I hypothesised that spatio-temporal information is particularly important for accurate developmental event detection given that many events during embryonic development of *L. stagnalis* have a distinct physiological or behavioural component, such as the onset of cardiac function or the attachment of the foot for crawling. Two-dimensional models are commonplace in developmental biology, particularly for classification tasks, such as the identification of developmental stages (Liu et al., 2021; Pond et al., 2021), the detection of deformations in embryonic phenotypes (Capek et al., 2023) and the prediction of developmental time (Jones et al., 2023; Toulany et al., 2023). Yet, the potential for greater accuracy from the application of a

3D CNN to quantifying behavioural and physiological developmental events was previously untested. I found significant increases in classification accuracy with the addition of temporal information i.e. using video instead of still images. Studies that incorporate deep learning models to automate classification of imaging data often use high-throughput imaging platforms to record developing embryos (e.g. Liu et al., 2019; Jones et al., 2023; Čapek et al., 2023). Therefore the integration of temporal information into classification tasks, such as in the use of Dev-ResNet, would enable improved classification accuracy with limited additional investment of time.

Whilst not directly applied to the experimental biology chapters of my thesis, the approach in Chapter 3 (Ibbini et al., 2024) for visualising embryonic development in high-dimensional space using Triplet loss and dimensionality reduction, represents a novel technique with which to achieve characterizations with greater granularity than possible with traditional approaches. Historically, the continuous process of embryonic development has been characterised using developmental events as standardised 'time–stamps' (Gould 1977; Spicer and Rundle 2006). However, visualising the differences between developmental events and how they relate to the continuous process of development remains a significant challenge, requiring an approach to capture adequate information throughout the embryonic period and distil it down to a human interpretable level, such as a two-dimensional mapping. I achieved this in Chapter 3 (Ibbini et al., 2024) by re-training Dev-ResNet using Triplet loss, a method that optimises a model to recognise the differences or similarities of samples depending on their class, which in this thesis were developmental events. Processing the final layer of this model using dimensionality

reduction algorithms including principal component analysis (PCA) enabled the visualisation of the entire embryonic period, as seen by the neural network, in a two-dimensional space. A similar, but unsupervised approach, was proposed by Toulany et al., (2023), where twin neural networks were used to identify similarities between a reference image timelapse of zebrafish embryonic development and new candidate timelapses, such as those of embryos exposed to different temperatures. As far as I am aware, these approaches are some of the first to use deep learning to effectively characterise embryonic development in a human interpretable visualisation, and could provide a more granular method in which to characterise development in experimental biology beyond what can be achieved with traditional approaches such as developmental events.

Overall, the combination of these deep learning approaches enabled the extraction of longitudinal data regarding some of the most commonly used phenotypic traits in the field of fluctuating temperatures, growth and development. The deep learning approaches presented in Chapter 4, though not novel in and of themselves, represent a successful demonstration of deep learning in experimental biology with minimal training data, a common scenario for many biological researchers. Moreover, the novel 3D CNN presented in Chapter 3 achieved a significant improvement for classification tasks over commonly used 2D models, within the context of animal development. The development and application of deep learning models here demonstrate the efficacy of this form of bioimage analysis for experimental biology, and in this thesis, facilitated the investigation of how developing ectotherms respond beyond the single endpoints commonly used in the literature.

6.3 | Developmental responses to fluctuating temperatures in *Lymnaea stagnalis*

The study of developmental responses to natural temperature variability has spanned a considerable period of time, with researchers in the early 1900s assessing the effects of fluctuating temperatures on economically important insect species (reviewed in Uvarov 1931), through to more recent theoretical and modelling studies that have integrated the findings of empirical studies into predictions around species-level responses to climate change (Deutsch et al., 2008; Vasseur et al., 2014; Levy et al., 2019). A central theme in this body of work is the usage of responses to constant temperatures to inform understanding of fluctuating temperatures found in the environment. The most prominent framework for achieving this involves the use of thermal performance curves (TPCs), and though its limitations have received considerable debate (reviewed in Sinclair et al., 2016), few alternatives exist and TPCs are still frequently used in empirical and theoretical studies today (e.g. Marshall et al., 2021; von Schmalensee et al., 2022). Perhaps the most common use of thermal performance curves in predicting responses to fluctuating temperatures involves the use of rate summation, whereby a cumulative sum of incremental rates, as described by a TPC constructed from constant temperature responses, is used to quantify a given developmental outcome for the duration of exposure (Worner 1992; Niehaus et al., 2012). Consequently a key aim in my thesis was to investigate the validity of rate summation, and this involved testing the validity of key assumptions of the approach, namely 1) that the relationship between performance and

temperature remains constant over the duration of exposure or throughout development (Niehaus et al., 2012; Kingsolver and Woods 2016), 2) that chronic and acute thermal exposure have the same physiological effects (Niehaus et al., 2012; Kingsolver et al., 2015; Sinclair et al., 2016), and 3) that thermal experience does not alter the response to the current temperature (Kingsolver and Woods 2016). While a number of empirical studies have focused on this problem (Niehaus et al., 2012; Arrighi et al., 2013; Kingsolver et al., 2015; Khelifa et al., 2019; Marshall et al., 2021), none have had the longitudinal data required to assess the validity of these key assumptions of the rate summation approach. The application of automated imaging and deep learning approaches presented in this thesis enabled the measurement of embryonic responses throughout development, and thus enabled me to test the validity of these key assumptions within an empirical context. Below I will discuss the contribution of my thesis to the literature in terms of the key assumptions of the rate summation approach outlined above.

Assumption 1: Thermal performance does not remain static over the course of development

Thermal performance curves used in rate summation are constructed on the basis of overall performance, for example, developmental rate quantified as the time from 1st cell division to hatch (Keen 1978; Taylor and Shields 1990; Khelifa et al., 2019). Thus, a key assumption of developmental studies that employ a rate summation approach is that performance in a given trait does not change during development, because a single TPC used to describe overall trait performance (e.g. rate from 1st cell division to hatch) is used to predict acute responses (e.g. hourly developmental rate) (Worner 1992; Niehaus et al., 2012; Kingsolver and Woods 2016). In Chapter 4, I demonstrated that this is not the case in embryos of *L. stagnalis* incubated to a

wide range of constant temperatures, where there was significant reductions in key parameters of TPCs, namely thermal optimum (T_{opt}), critical upper limit (CT_{max}) and thermal breadth between developmental events.

Differences in thermal performance between developmental stages have been recorded across a number of taxa in thermal tolerance studies (e.g. Storch et al., 2011; Komoroske et al., 2014; Truebano et al., 2018; Hall and Warner 2019; Ruthsatz et al., 2022), though individuals are typically acclimated to a control temperature up until the stage at which the thermal assay is carried out to remove thermal history as a confounding factor. However, these studies are not directly applicable as an explanation for inaccuracies associated with rate summation found in empirical studies of fluctuating temperatures, where individuals are incubated to chronic temperatures for the duration of the developmental period (e.g. Niehaus et al., 2012; Kingsolver et al., 2015; Khelifa et al., 2019). Consequently, a primary focus of Chapter 4 was to incubate embryos to the same chronic thermal assays used in empirical studies of fluctuating temperatures, and assess thermal performance at key developmental events during embryonic development. As far as I am aware, these data are amongst the first to describe thermal performance in a longitudinal manner for embryos incubated to chronic temperatures for the duration of development. Furthermore, the contraction in thermal performance observed across developmental periods mirrors the theoretical findings of Kingsolver and Woods (2016), where a mechanistic model incorporating the costs of HSP synthesis on growth rate predicted a similar contraction in upper thermal performance with increased duration of exposure. Development can encompass large changes across behaviour, morphology and physiology, and

so it would be unsurprising that thermal sensitivity would differ between developmental stages. Furthermore, the impact of variable temperatures, particularly if they are stochastic as they are in the environment, could alter the thermal sensitivity of later developmental stages depending on the thermal experience of earlier stages. Consequently, developing animals could follow a myriad of different pathways depending on their thermal experience.

The limitations associated with this key assumption of rate summation highlights that perhaps we are attempting to use rate summation to predict developmental responses on physiological time scales that are too long. Given that rate summation inherently assumes no changes in thermal performance during exposure (Tanigoshi and Logan 1979; Worner 1992; Niehaus et al., 2012), then application of rate summation to individual developmental periods or stages might provide more accurate results, given that the discrepancies caused by acclimation or thermal stress might be less apparent on shorter physiological time-scales. For example, accurate predictions of developmental rate were achieved for some early life stages of the toad, *Bombina orientalis* (Arrighi et al., 2013) and in the armyworm, *Pseudaletia unipuncta* (Taylor and Shields 1990). Similarly, accurate predictions for some early developmental events in terms of developmental and growth rate were achieved in Chapter 5 of this thesis, and the absolute error between predicted and observed measures was comparatively small relative to late developmental events. However, a challenge in using a developmental period- or stage-specific approach is in the approach for constructing thermal performance curves for each developmental period. If we use the same approach in this thesis, where embryos are incubated to the same constant temperature until hatch, then there is a confounding factor of thermal

experience in TPCs for late developmental periods. Alternatively, one could begin exposure of embryos to constant treatments immediately after the onset of a developmental event or stage as is common in thermal tolerance literature (e.g. Storch et al., 2011; Komoroske et al., 2014; Ruthsatz et al., 2022). Of course, acute thermal assays, such as ramping assays could also be conducted, however, these would likely not allow the measurement of slower responding traits such as growth and developmental rate between stages. Regardless, by partitioning the entirety of embryonic development into smaller periods and applying the rate summation approach to each period individually, more accurate predictions of whole-developmental responses might be achievable.

Assumption 2: Thermal performance to chronic exposure is not equivalent to acute exposure

In the empirical and theoretical literature covering biological responses to fluctuating temperatures, it has long been assumed that performance measured under chronic exposure is equivalent to that of acute exposure (Niehaus et al., 2012; Kingsolver et al,. 2015). Though empirical data disproving the validity of this assumption have been recorded as early as the 1920s (e.g. Cook 1927), it has not received attention until relatively recently (Niehaus et al., 2012; Kingsolver and Woods 2016). In Chapter 5, I demonstrated that *L. stagnalis* embryos could still continue development and hatch despite acute exposure to chronically lethal temperatures, and that hourly growth rates measured in fluctuating temperatures far exceeded predicted growth rate on the basis of rate summation. These data clearly indicate that responses derived from chronic exposure are likely to be ill-suited to predict responses to acute exposure in developing ectotherms.

Empirical studies quantifying the effects of thermal fluctuations on short-term responses (e.g. hourly feeding rate) observe similar impacts of the nonlinearity inherent within TPCs as on long-term responses (Vajedsamiei et al., 2021), though the range of operative temperatures is far greater than what individuals can tolerate over longer duration exposures. These differences in responses depending on duration of exposure are commonplace in the thermal tolerance literature (Rezende et al., 2014; Jorgensen et al., 2021), and have been integrated into theoretical frameworks such as thermal-death-time (TDT) curves (Rezende et al,. 2014). The TDT approach exposes animals to static thermal assays to generate tolerance times for a range of temperatures (Rezende et al., 2011). Log-transformed tolerance times are then plotted against exposure temperature to reveal a linear relationship between temperature and thermal tolerance, referred to as the TDT curve (Rezende et al., 2011). The integration of tolerance times into this approach enables the identification of trade-offs of tolerance to chronic vs acute exposure, such as in different life stages of an intertidal gastropod (Truebano et al., 2018). Creating multi-dimensional landscapes similar to TDT curves for other developmental responses, such as growth, would undoubtedly improve visualisation of how different characteristics of FT regimes affect development. Indeed, researchers have long attempted to visualise such a multi-dimensional function in the context of fluctuating temperatures (e.g. Cook 1927; Messenger and Flitters 1959; Kingsolver and Woods 2016; Cavieres et al., 2018)(Fig. 6.1) but they have lacked sufficient empirical data to accurately characterise this landscape (e.g. Fig. 6.1A). Theoretical studies have warned that such empirical data would not be suited for use with rate summation, because rate summing TPCs derived from fluctuating regimes would

underestimate actual responses because the same effects of Jensen's inequality would apply to these TPCs (Worner 1992). Nevertheless, integrating empirical data into a multi-dimensional landscape, such as one relating mean temperature, amplitude of natural thermal regimes and biological rate of biological response(Fig. 6.1B), would undoubtedly improve our understanding of how developing ectotherms respond to naturally fluctuating temperatures, beyond the limited number of treatments used in empirical studies to date (e.g. Fig. 6.1A).

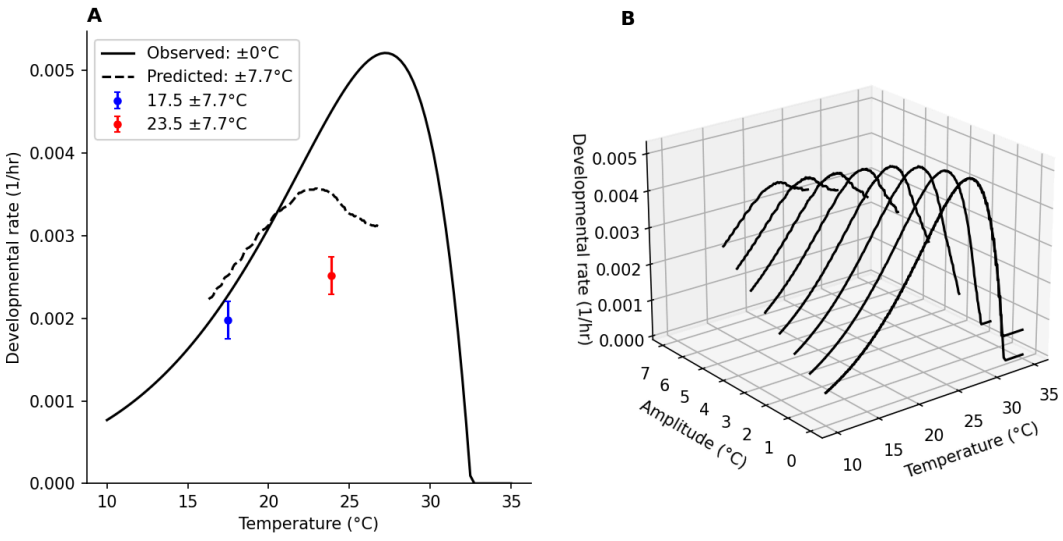

Figure 6.1. Relationship between temperature and developmental rate for *Lymnaea stagnalis* embryos (A), where the solid line represents a thermal performance curve (TPC) fit to responses observed at constant temperatures (Chapter 4, Fig. 4.1), whilst the dashed line represents predicted developmental rate derived from rate summation (Chapter 5, Fig. 5.2). Developmental rates observed at two fluctuating temperature regimes (17.7 ±7.7°C and 23.5 ±7.7°C) are also overlaid (A) (Chapter 5, Fig. 5.3). Secondly, a 3D landscape illustrating the predicted effect of both amplitude and average temperature on developmental rate (B), where data for amplitudes > 0°C are derived from rate summation of the TPC at an amplitude of 0°C, i.e. constant temperature. In both (A) and (B), rate summation is only performed where the range of temperatures in the fluctuating temperature regime does not exceed the upper and lower critical limits of the TPC.

Assumption 3: Thermal experience does alter the response to the current temperature; the importance of testing for active plasticity

Thermal experience can alter the response to the current temperature through plastic responses, whether they be active, i.e. involving a biochemical or physiological response to a stimulus, or a 'passive', biophysical response (Schulte et al., 2011). Yet, rate summation and its derivatives assume only passive plasticity (Schulte et al., 2011), since temperature dependence at a given temperature is modelled by a single TPC for the entirety of development or an exposure period (Worner 1992). However, active plastic developmental responses have been recorded in empirical studies to a range of environmental drivers, including salinity (Tills et al., 2010), temperature (Pottier et al., 2022) and dissolved oxygen (Ruhr et al., 2019).

In the present thesis, *L. stagnalis* embryos exhibited markedly different developmental trajectories in terms of growth under fluctuating temperature incubation (Chapter 5), and thus applying a rate summation approach that assumes no changes in thermal performance during development will inevitably have limitations in its accuracy of predictions. von Schmalensee et al., (2021) found that rate summation could effectively predict developmental times in embryos of the butterfly *Pieres napi* incubated to field conditions, and they concluded that *Pieres napi* is unaffected by thermal fluctuations during development. Such passive plasticity could therefore be accurately captured by a rate summation approach (Schulte et al., 2011), but longitudinal data describing these responses would be required to assess whether predicted developmental rates match empirical responses over the entire exposure period and not just for developmental endpoints such as hatch time.

Altered developmental trajectories exhibited by *L. stagnalis* embryos in response to fluctuating temperature incubation indicate an actively plastic response (Schulte et al., 2011). Though, this response is potentially mal-adaptive given that embryos hatched at a smaller size in comparison to constant temperature incubation at the same mean temperature (Chapter 5), and increased size is typically linked with greater fitness among a population or species (Kingsolver and Huey 2008). Future work that assesses developmental responses in different populations of this species would prove useful in identifying whether these plastic responses are specific to this population or representative of the species as a whole. If they are representative of responses found in other populations, then the acquisition of longitudinal data regarding developmental responses to different fluctuating thermal regimes would enable identification of the relative impact of different extreme temperature events on this species, such as how development is altered under conditions typical of a cold snap or a heat wave. Given that climate change is predicted to increase the incidence of extreme climatic events (Easterling et al., 2000; Meehl and Tebaldi 2004; Rummukainen 2012), then such empirical data would prove crucial in assessing the vulnerability of this species' earliest life stages.

6.4 | Conclusion

Thermal performance curves have been used extensively in approaches to understand and predict the responses of ectotherms to fluctuating temperatures, with rate summation remaining perhaps the most common approach (Worner 1992; Niehaus et al., 2012; Sinclair et al., 2016). From the findings of this thesis, it seems that rate summation of thermal performance curves is poorly positioned to provide accurate predictions of developmental responses, whether to specific developmental events or whole developmental responses, such as time from 1st cell division to hatch. Rate summation however, could still prove to be a useful scaffold at short physiological time scales (e.g. between developmental events), upon which other theoretical approaches could be used, such as time-dependent effects (Kingsolver and Woods 2016) and relaxation responses (Roitberg and Mangel 2016). The integration of high-throughput imaging technologies with automated bioimage analyses enabled the extraction of granular longitudinal data, facilitating the assessment of key assumptions of the rate summation approach that has not been achieved in empirical studies previously. Thus, the application of these methodologies could prove to be a viable vehicle to provide the empirical data necessary to gather a more complete understanding of ectotherm responses to fluctuating temperatures. Regardless, further theoretical and empirical work is required before we are able to accurately capture the responses of developing ectotherms to fluctuating temperatures, and therefore accurately predict species vulnerability to future climatic scenarios, reliably for all life stages.

Appendix

Sections of this Appendix have been published in the following papers:

Ibbini, Z., Spicer, J.I., Truebano, M., Bishop, J. and Tills, O., 2022. HeartCV: a tool for transferrable, automated measurement of heart rate and heart rate variability in transparent animals. Journal of Experimental Biology, 225(19), p.jeb244729.

Ibbini, Z., Truebano, M., Spicer, J.I., McCoy, J. and Tills, O., 2024. Dev-ResNet: automated developmental event detection using deep learning. Journal of Experimental Biology, 227(10).

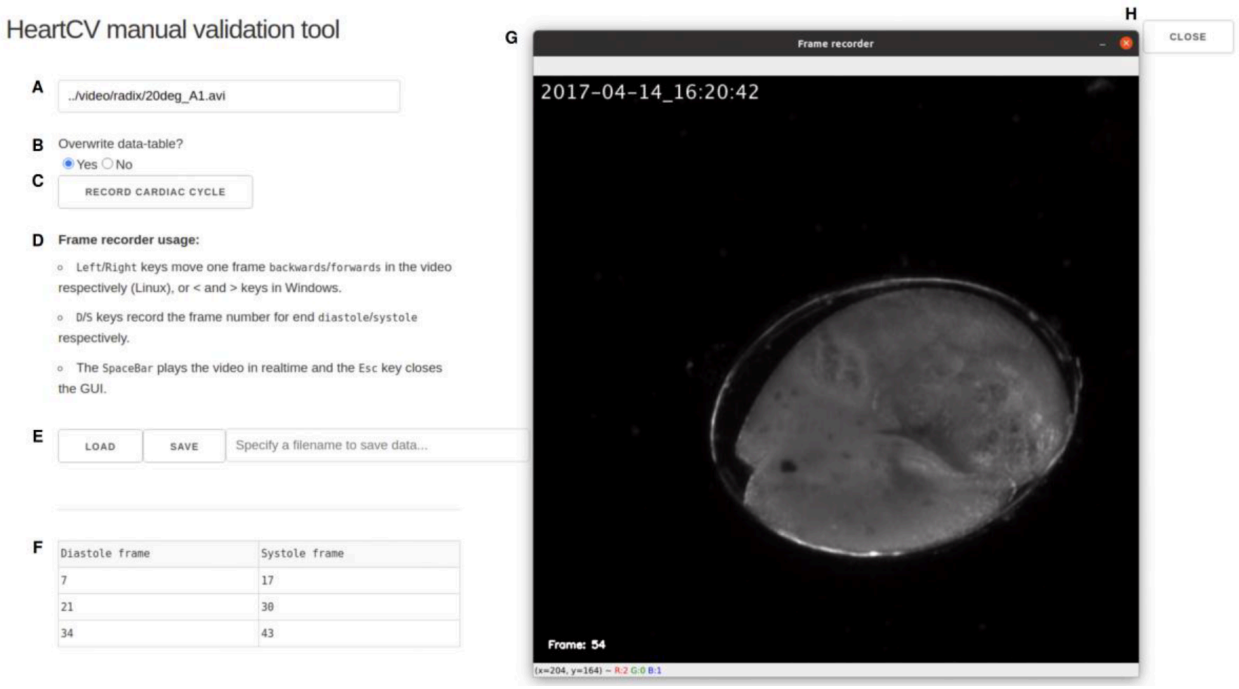

Figure A1. User interface for manual quantification of heart rate and IBI measures for comparison with measures derived from automated quantification. Users can specify individual videos to validate via a file input (A) and whether over-write existing data through the use of toggle buttons (e.g. when validating multiple videos in one session) (B). The frames at which a given cardiac event occurs can be recorded through the use of an interactive window (C and G) which is responsive to keyboard presses (D). These results are then stored in a table (F) and can be exported to CSV format (E).

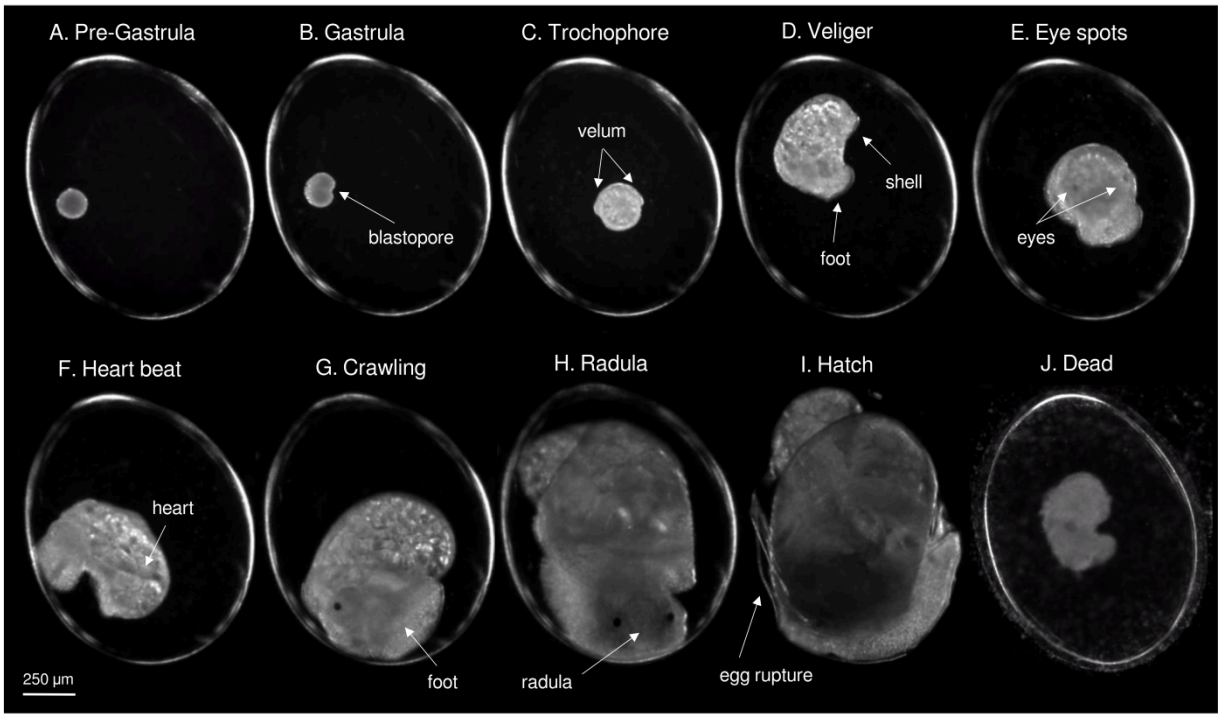

Figure A2. Developmental events used in this study are shown (A-D), with key characteristics of each event indicated by the white arrows and text

Table A1. Descriptions of developmental events examined for *Lymnaea stagnalis,* after Meshcheryakov (1990) and Smirthwaite et al., (2007), including event timings observed at a reference temperature of 20°C. Pre-Gastrula timings are excluded because we use this descriptor for the developmental period prior to gastrulation and not as an event itself, and timings of mortality ('Dead') are omitted because no mortality was observed at a reference temperature of 20°C.

Developmental event	Description of event	Event time (hours) at 20°C (mean ± sd)
Pre-Gastrula	Encompasses cell divisions and flattening of embryo at poles.	N/A
Gastrula	Clear visible formation of blastospore and cilia driven rotation is now observable.	49.2 ± 3.9
Trochophore	Liver cells form a helmet-like velum over the embryo.	89.0 ± 5.0
Veliger	Distinct formation of the two poles of the embryo into the shell and the foot, with each pole now being clearly distuingishable.	122.5 ± 9.5
Eye spots	Pigmentation of the eye spots can be observed on the head, and they continue to darken as the embryo develops.	143.9 ± 4.0
Heart beat	A rhythmic beat can be observed in the heart situated in the shell of the embryo. The heart gradually moves toward the head of the embryo as it grows.	150.9 ± 3.2
Crawling	The foot of the embryo attaches to the egg capsule, enabling it to crawl.	176.8 ± 10.3

Radula	The radula, located in the head between the eye spots, begins to suck in capsular fluid.	239.2 ± 3.2
Hatch	Rupture of the egg capsule, followed by the emergence of the embryo.	295.4 ± 3.7
Dead	Cessation of development and movement, often accompanied by rapid growth of microbial organisms in the egg.	N/A

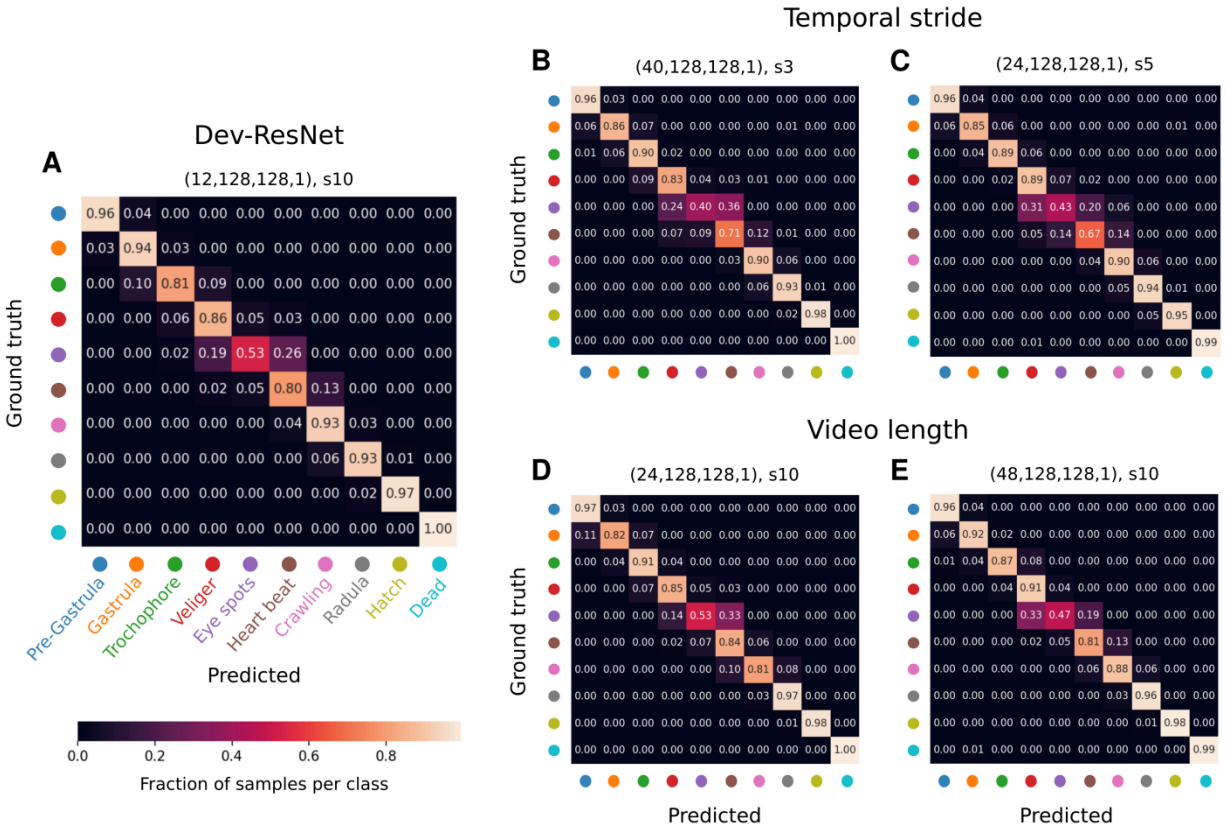

Figure A3. Confusion matrices comparing counts between ground truth and predicted developmental event labels on the testing dataset for Dev-ResNet trained with different temporal strides: 3 (B), 5 (C) and 10 frames (A), but also with different video lengths: 12 (A), 24 (D) and 48 frames (E) using the same temporal stride (10 frames). The

axis tick colours assigned to each event are described in (A). Input shape (length, width, height, channels) and temporal stride notation (e.g. s10 equals a stride of 10 frames) corresponding to each model are subtitles above each plot.

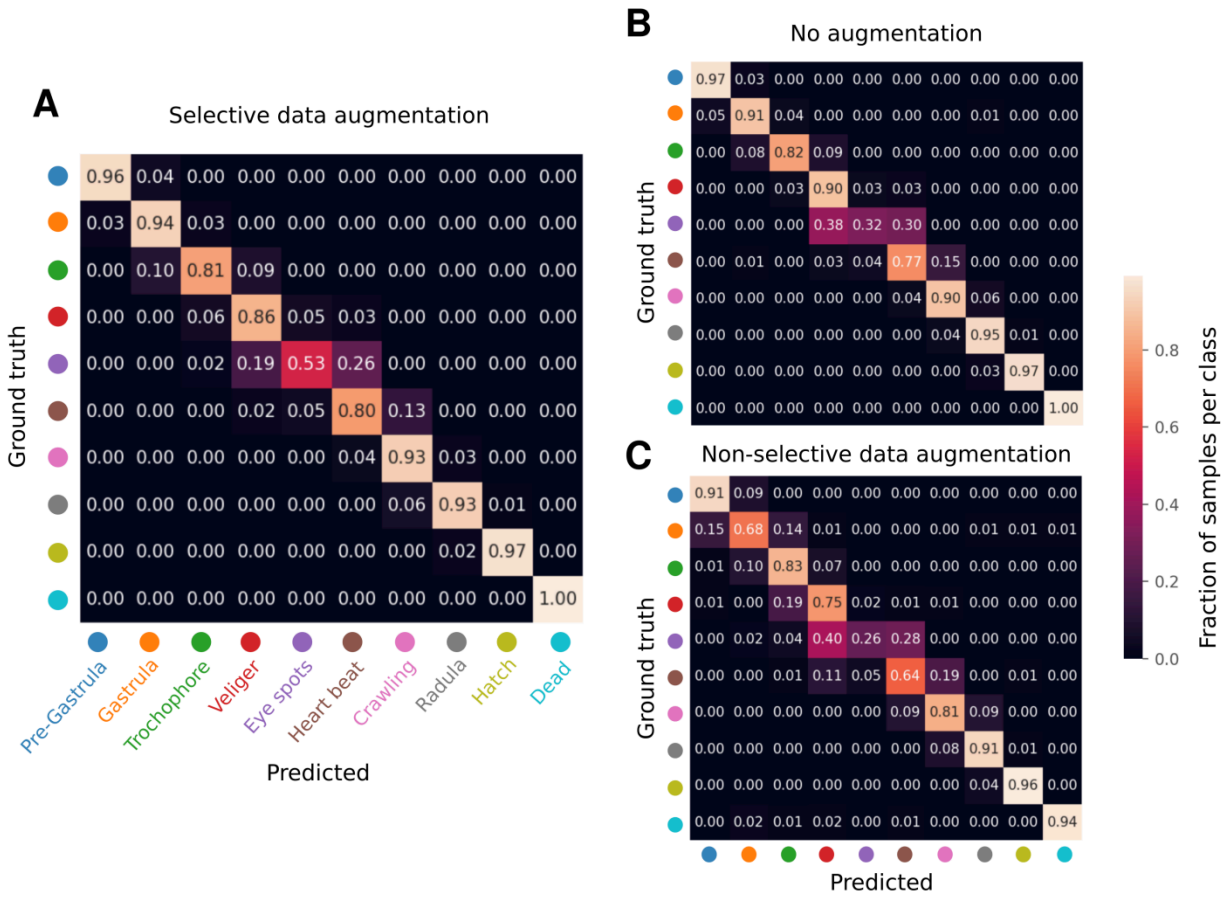

Figure A4. Confusion matrices comparing counts between ground truth and predicted developmental event labels on the testing dataset for Dev-ResNet trained with no visual augmentations (B), with non-selective augmentation, i.e. overall training samples increased but class imbalances remained (C) and with selective data augmentation, i.e. class imbalances were removed by applying augmentation to varying degrees depending on a given class' representation in the original data (A). The axis tick colours assigned to each event are described in (A).

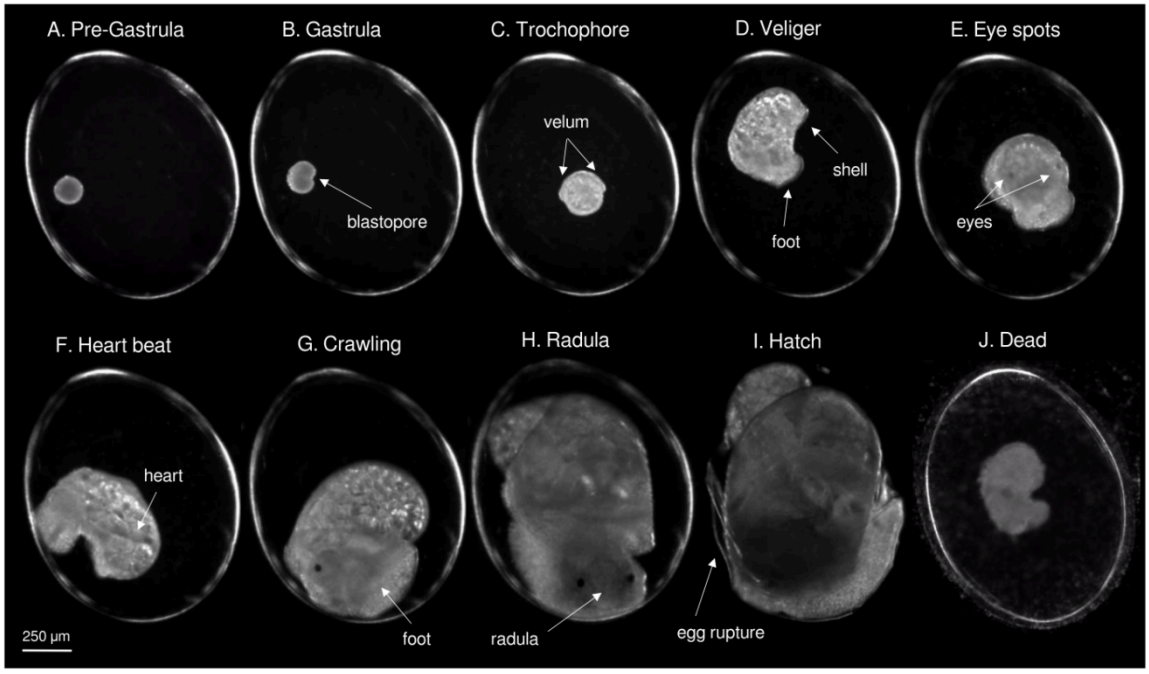

Figure A5. Developmental events used in this study are shown (A-D), with key characteristics of each event indicated by the white arrows and text.

Table A2. Comparisons of models fit to the developmental rates of *L. stagnalis* embryos at each developmental event. For each model, we report both the AIC and the differential AIC (Δi), which is the difference between a given model's AIC and the lowest AIC. The best-fitting model for each developmental event is denoted with bold font.

Developmental event	Model	AIC	Δi
Gastrula	Gaussian	-70.94	4.87
	Weibull	-72.26	3.55
	Rezende et al., (2019)	**-75.81**	**0.00**
	Thomas et al., (2017)	-71.50	4.31
Trochophore	Gaussian	-83.83	14.32
	Weibull	-94.02	4.13
	Rezende et al., (2019)	**-98.15**	**0.00**
	Thomas et al., (2017)	-96.21	1.94
Veliger	Gaussian	-74.45	18.84
	Weibull	-78.00	15.29
	Rezende et al., (2019)	**-93.29**	**0.00**
	Thomas et al., (2017)	-89.49	3.8
Eye spots	Gaussian	-64.01	32.36
	Weibull	-66.47	29.9
	Rezende et al., (2019)	**-96.37**	**0.00**
	Thomas et al., (2017)	-88.55	7.82
Heart beat	Gaussian	-67.50	28.75
	Weibull	-70.17	26.08
	Rezende et al., (2019)	**-96.25**	**0.00**
	Thomas et al., (2017)	-91.33	4.92
Crawling	Gaussian	-72.60	35.22
	Weibull	-75.99	31.83
	Rezende et al., (2019)	-107.46	0.36
	Thomas et al., (2017)	**-107.82**	**0.00**
Radula	Gaussian	-75.53	37.73
	Weibull	-79.15	34.11
	Rezende et al., (2019)	-110.14	3.12
	Thomas et al., (2017)	**-113.26**	**0.00**
Hatch	Gaussian	-82.17	26.56
	Weibull	-86.35	22.38
	Rezende et al., (2019)	**-108.73**	**0.00**
	Thomas et al., (2017)	-94.32	14.41

Table A3. Comparisons of models fit to the growth rates of *L. stagnalis* embryos at each developmental event. For each model, we report both the AIC and the differential AIC (Δi), which is the difference between a given model's AIC and the lowest AIC. The best-fitting model for each developmental event is denoted with bold font.

Developmental event	Model	AIC	Δi
Gastrula	**Gaussian**	**-130.64**	**0.00**
	Weibull	-129.34	1.30
	Rezende et al., (2019)	-128.64	2.00
	Thomas et al., (2017)	-127.20	3.44
Trochophore	Gaussian	-124.17	0.00
	Weibull	-122.8	1.37
	Rezende et al., (2019)	-122.93	1.23
	Thomas et al., (2017)	-121.21	2.95
Veliger	**Gaussian**	**-124.83**	**0.00**
	Weibull	-123.31	1.52
	Rezende et al., (2019)	-121.74	3.09
	Thomas et al., (2017)	-120.47	4.36
Eye spots	Gaussian	-101.77	23.03
	Weibull	-104.66	20.14
	Rezende et al., (2019)	**-124.8**	**0.00**
	Thomas et al., (2017)	-112.09	12.70
Heart beat	Gaussian	-99.21	20.85
	Weibull	-102.65	17.40
	Rezende et al., (2019)	**-120.05**	**0.00**
	Thomas et al., (2017)	-114.01	6.04
Crawling	Gaussian	-95.75	20.26
	Weibull	-98.71	17.30
	Rezende et al., (2019)	**-116.01**	**0.00**
	Thomas et al., (2017)	-113.19	2.82
Radula	Gaussian	-92.94	15.41
	Weibull	-95.27	13.08
	Rezende et al., (2019)	**-108.35**	**0.00**
	Thomas et al., (2017)	-101.97	6.38
Hatch	Gaussian	-95.23	13.91
	Weibull	-97.91	11.23
	Rezende et al., (2019)	**-109.14**	**0.00**
	Thomas et al., (2017)	-106.78	2.36

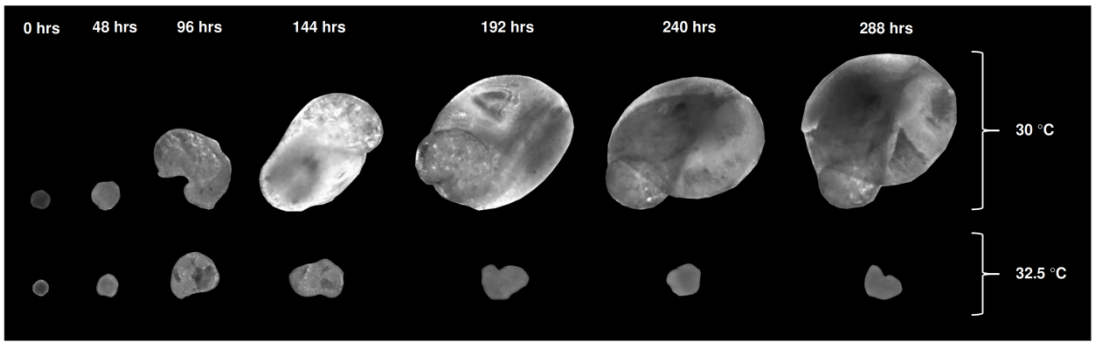

Figure A6. Developmental timeline for *L. stagnalis* in constant temperature treatments of 30 and 32.5C.

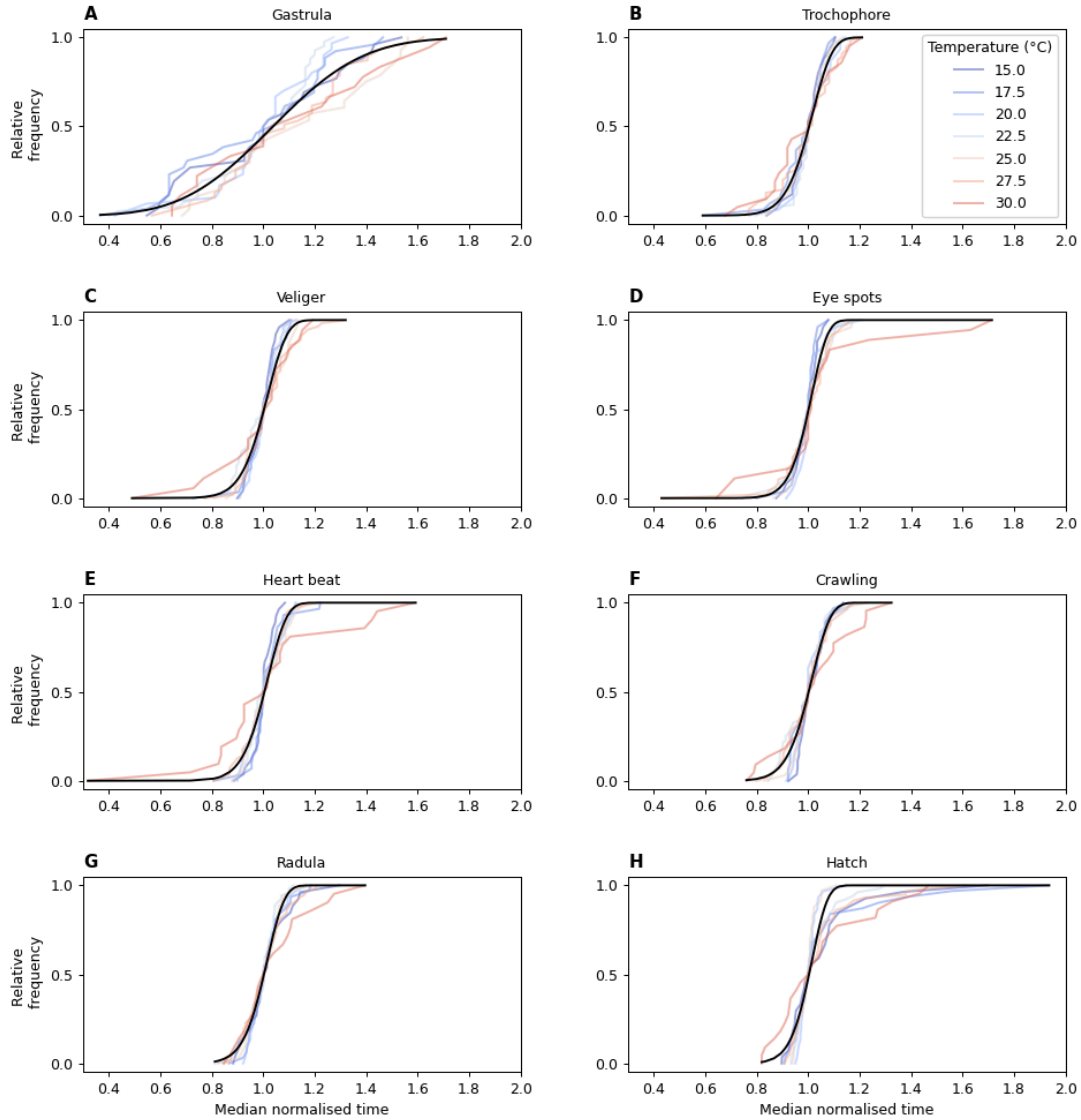

Figure A7. Median normalised cumulative distributions of developmental event timings in embryonic development of *Lymnaea stagnalis* across a range of chronic incubation temperatures. Solid lines represent a curve of best fit of the Weibull cumulative distribution function.

Table A4. Statistical comparisons of observed and predicted rates of growth and developmental rate in embyros of *Lymnaea stagnalis*.

Developmental event	Thermal regime (°C)	Developmental rate (hr^{-1})	Growth rate ($mm^2\ hr^{-1}$)
Gastrula	17.7 ± 7.7	$t_{21} = 3.75, p < 0.01$	$t_{21} = 2.88, p < 0.001$
	23.5 ± 7.7	$t_{11} = 6.09, p < 0.001$	$t_{11} = 7.39, p < 0.05$
Trochophore	17.7 ± 7.7	$t_{21} = 6.51, p < 0.0001$	$t_{21} = -4.54, p < 0.001$
	23.5 ± 7.7	$t_{11} = 1.90, p < 0.05$	$t_{11} = -8.09, p < 0.05$
Veliger	17.7 ± 7.7	$t_{21} = 7.09, p < 0.0001$	$t_{21} = -5.49, p < 0.001$
	23.5 ± 7.7	$t_{11} = 1.46, p = 0.1936$	$t_{11} = -3.83, p < 0.05$
Eye spots	17.7 ± 7.7	$t_{21} = 5.37, p < 0.0001$	$t_{21} = -8.52, p < 0.001$
	23.5 ± 7.7	$t_{11} = 8.98, p = 0.3903$	$t_{11} = -5.35, p < 0.05$
Heart beat	17.7 ± 7.7	$t_{21} = 8.22, p < 0.0001$	$t_{21} = -4.50, p < 0.001$
	23.5 ± 7.7	$t_{11} = 3.17, p < 0.01$	$t_{11} = -5.96, p < 0.05$
Crawling	17.7 ± 7.7	$t_{21} = 6.97, p < 0.0001$	$t_{21} = -3.54, p < 0.001$
	23.5 ± 7.7	$t_{11} = 2.87, p < 0.01$	$t_{11} = -3.43, p < 0.05$
Radula	17.7 ± 7.7	$t_{21} = 1.06, p < 0.0001$	$t_{21} = -3.35, p < 0.001$
	23.5 ± 7.7	$t_{11} = 9.33, p < 0.0001$	$t_{11} = -1.07, p < 0.05$
Hatch	17.7 ± 7.7	$t_{21} = 8.41, p < 0.0001$	$t_{21} = -7.38, p < 0.001$
	23.5 ± 7.7	$t_{11} = 7.90, p < 0.0001$	$t_{11} = -6.83, p < 0.05$

References

Abram, P.K. *et al.* (2016) 'Behavioural effects of temperature on ectothermic animals: Unifying thermal physiology and behavioural plasticity', *Biological Reviews*, 92(4), pp. 1859–1876. doi:10.1111/brv.12312.

Amorim, J. *et al.* (2019) 'Lymnaea stagnalis as a freshwater model invertebrate for ecotoxicological studies', *Science of The Total Environment*, 669, pp. 11–28. doi:10.1016/j.scitotenv.2019.03.035.

Andronikov, V.B. (1975) 'Heat resistance of gametes of marine invertebrates in relation to temperature conditions under which the species exist', *Marine Biology*, 30(1), pp. 1–11. doi:10.1007/bf00393747.

Anger, K. *et al.* (2003) 'Larval and early juvenile development of Paralomis granulosa reared at different temperatures: Tolerance of cold and food limitation in a lithodid crab from high latitudes', *Marine Ecology Progress Series*, 253, pp. 243–251. doi:10.3354/meps253243.

Angilletta, M.J. *et al.* (2003) 'Tradeoffs and the evolution of thermal reaction norms', *Trends in Ecology & Evolution*, 18(5), pp. 234–240. doi:10.1016/s0169-5347(03)00087-9.

Angilletta, M.J. (2006) 'Estimating and comparing thermal performance curves', *Journal of Thermal Biology*, 31(7), pp. 541–545. doi:10.1016/j.jtherbio.2006.06.002.

Angilletta, M.J. (2014) *Thermal adaptation: A theoretical and empirical synthesis*. Oxford: Oxford University Press.

Arrighi, J.M. *et al.* (2013) 'Daily temperature fluctuations unpredictably influence developmental rate and morphology at a critical early larval stage in a frog', *BMC Ecology*, 13(1). doi:10.1186/1472-6785-13-18.

Atli, G. and Grosell, M. (2016) 'Characterization and response of antioxidant systems in the tissues of the freshwater pond snail (Lymnaea stagnalis) during acute copper exposure', *Aquatic Toxicology*, 176, pp. 38–44. doi:10.1016/j.aquatox.2016.04.007.

Audira, G. *et al.* (2020) 'Systematical exploration of the common solvent toxicity at whole organism level by behavioral phenomics in adult zebrafish', *Environmental Pollution*, 266, p. 115239. doi:10.1016/j.envpol.2020.115239.

Audira, G. *et al.* (2021) 'Phenomics approach to investigate behavioral toxicity of environmental or occupational toxicants in adult zebrafish (*danio rerio*)', *Current Protocols*, 1(8). doi:10.1002/cpz1.223.

Baden, T. *et al.* (2015) 'Open Labware: 3-D printing your own lab equipment', *PLOS Biology*, 13(3). doi:10.1371/journal.pbio.1002086.

Bandow, C. and Weltje, L. (2012) 'Development of an embryo toxicity test with the pond snail Lymnaea stagnalis using the model substance tributyltin and common solvents', *Science of The Total Environment*, 435–436, pp. 90–95. doi:10.1016/j.scitotenv.2012.07.005.

Bartheld, J.L., Artacho, P. and Bacigalupe, L. (2017) 'Thermal performance curves under daily thermal fluctuation: A study in helmeted water toad tadpoles', *Journal of Thermal Biology*, 70, pp. 80–85. doi:10.1016/j.jtherbio.2017.09.008.

Battisti, David.S. and Naylor, R.L. (2009) 'Historical warnings of future food insecurity with unprecedented seasonal heat', *Science*, 323(5911), pp. 240–244. doi:10.1126/science.1164363.

Bernhardt, J.R. *et al.* (2018) 'Nonlinear averaging of thermal experience predicts population growth rates in a thermally variable environment', *Proceedings of the Royal Society B: Biological Sciences*, 285(1886), p. 20181076. doi:10.1098/rspb.2018.1076.

Bilder, R.M. *et al.* (2009) 'Phenomics: The Systematic Study of phenotypes on a genome-wide scale', *Neuroscience*, 164(1), pp. 30–42. doi:10.1016/j.neuroscience.2009.01.027.

Bowler, K. and Terblanche, J.S. (2008) 'Insect thermal tolerance: What is the role of ontogeny, ageing and senescence?', *Biological Reviews*, 83(3), pp. 339–355. doi:10.1111/j.1469-185x.2008.00046.x.

Bozinovic, F. *et al.* (2011) 'The mean and variance of environmental temperature interact to determine physiological tolerance and fitness', *Physiological and Biochemical Zoology*, 84(6), pp. 543–552. doi:10.1086/662551.

Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools.

Brainerd, E.L. and Hale, M.E. (2006) 'In vivo and functional imaging in Developmental Physiology', *Comparative Developmental Physiology*, pp. 21–40. doi:10.1093/oso/9780195168594.003.0002.

Bubliy, O.A. and Loeschcke, V. (2002) 'Effect of low stressful temperature on genetic variation of five quantitative traits in drosophila melanogaster', *Heredity*, 89(1), pp. 70–75. doi:10.1038/sj.hdy.6800104.

Burggren, W. (2021) ‘Developmental physiology: Grand challenges’, *Frontiers in Physiology*, 12. doi:10.3389/fphys.2021.706061.

Burggren, W.W. (2021) ‘Putting the August Krogh principle to work in Developmental Physiology’, *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 252, p. 110825. doi:10.1016/j.cbpa.2020.110825.

Burggren, W.W. and Mueller, C.A. (2015) ‘Developmental critical windows and sensitive periods as three-dimensional constructs in time and space’, *Physiological and Biochemical Zoology*, 88(2), pp. 91–102. doi:10.1086/679906.

Byrne, R.A. *et al.* (2009) ‘Embryonic rotational behaviour in the pond snail Lymnaea stagnalis: Influences of environmental oxygen and development stage’, *Zoology*, 112(6), pp. 471–477. doi:10.1016/j.zool.2009.03.001.

Čapek, D. *et al.* (2023) ‘EmbryoNet: Using deep learning to link embryonic phenotypes to signaling pathways’, *Nature Methods*, 20(6), pp. 815–823. doi:10.1038/s41592-023-01873-4.

Cardona, A. and Tomancak, P. (2012) ‘Current challenges in open-source Bioimage Informatics’, *Nature Methods*, 9(7), pp. 661–665. doi:10.1038/nmeth.2082.

Cavieres, G. *et al.* (2018) ‘Fluctuating thermal environments and time-dependent effects on fruit fly egg-hatching performance’, *Ecology and Evolution*, 8(14), pp. 7014–7021. doi:10.1002/ece3.4220.

Chagas, A.M. *et al.* (2017) *The 100 € lab: A 3-D printable open source platform for fluorescence microscopy, optogenetics and accurate temperature control during behaviour of zebrafish,drosophilaandC. elegans* [Preprint]. doi:10.1101/122812.

Chan, P.K., Lin, C.C. and Cheng, S.H. (2009) 'Noninvasive technique for measurement of heartbeat regularity in zebrafish (danio rerio) embryos', *BMC Biotechnology*, 9(1). doi:10.1186/1472-6750-9-11.

Chen, T.-J. *et al.* (2019) 'Using deep learning with large dataset of microscope images to develop an automated embryo grading system', *Fertility & Reproduction*, 01(01), pp. 51–56. doi:10.1142/s2661318219500051.

Chevin, L.-M., Lande, R. and Mace, G.M. (2010) 'Adaptation, plasticity, and extinction in a changing environment: Towards a predictive theory', *PLoS Biology*, 8(4). doi:10.1371/journal.pbio.1000357.

Chollet, F. (2017) 'Xception: Deep learning with depthwise separable convolutions', *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [Preprint]. doi:10.1109/cvpr.2017.195.

Cleveland, W.S. (1979) 'Robust locally weighted regression and smoothing scatterplots', *Journal of the American Statistical Association*, 74(368), p. 829. doi:10.2307/2286407.

Colinet, H. *et al.* (2015) 'Insects in fluctuating thermal environments', *Annual Review of Entomology*, 60(1), pp. 123–140. doi:10.1146/annurev-ento-010814-021017.

Colmorgen, M. and Paul, R.J. (1995) 'Imaging of physiological functions in transparent animals (Agonus cataphractus, Daphnia Magna, pholcus phalangioides) by video microscopy and Digital Image Processing', *Comparative Biochemistry and Physiology Part A: Physiology*, 111(4), pp. 583–595. doi:10.1016/0300-9629(95)00059-g.

Coma, R. *et al.* (2009) 'Global warming-enhanced stratification and mass mortality events in the Mediterranean', *Proceedings of the National Academy of Sciences*, 106(15), pp. 6176–6181. doi:10.1073/pnas.0805801106.

Conte, F.P. (1984) 'Structure and function of the crustacean larval salt gland', *International Review of Cytology*, pp. 45–106. doi:10.1016/s0074-7696(08)61314-5.

Cook, W.C. (1927) 'Some effects of alternating temperatures on the growth and metabolism of cutworm larvae1', *Journal of Economic Entomology*, 20(6), pp. 769–782. doi:10.1093/jee/20.6.769.

Dahlke, F.T. *et al.* (2020) 'Thermal bottlenecks in the life cycle define climate vulnerability of fish', *Science*, 369(6499), pp. 65–70. doi:10.1126/science.aaz3658.

Dallwitz, R. (1984) 'The influence of constant and fluctuating temperatures on development rate and survival of pupae of the Australian sheep blowfly *lucilia cuprina*', *Entomologia Experimentalis et Applicata*, 36(1), pp. 89–95. doi:10.1111/j.1570-7458.1984.tb03412.x.

Dantzler, W. H., (1997). Handbook of physiology: a critical, comprehensive presentation of physiological knowledge and concepts. Oxford University Press.

Darwin, C. (1859). On the Origin of Species, or the Preservation of Favoured Races in the Struggle for Life. London: John Murray.

Davidson, B. (2007) 'Ciona intestinalis as a model for cardiac development', *Seminars in Cell & Developmental Biology*, 18(1), pp. 16–26. doi:10.1016/j.semcdb.2006.12.007.

Davis, B.D. (1949) 'The isolation of biochemically deficient mutants of bacteria by means of penicillin', *Proceedings of the National Academy of Sciences*, 35(1), pp. 1–10. doi:10.1073/pnas.35.1.1.

De Luca, E. *et al.* (2014) 'ZebraBeat: A flexible platform for the analysis of the cardiac rate in zebrafish embryos', *Scientific Reports*, 4(1). doi:10.1038/srep04898.

Denny, M. (2017) 'The fallacy of the average: On the ubiquity, utility and continuing novelty of Jensen's inequality', *Journal of Experimental Biology*, 220(2), pp. 139–146. doi:10.1242/jeb.140368.

Deutsch, C.A. *et al.* (2008) 'Impacts of climate warming on terrestrial ectotherms across latitude', *Proceedings of the National Academy of Sciences*, 105(18), pp. 6668–6672. doi:10.1073/pnas.0709472105.

Dillon, M.E. and Woods, H.A. (2016) 'Introduction to the symposium: Beyond the mean: Biological impacts of changing patterns of temperature variation', *Integrative and Comparative Biology*, 56(1), pp. 11–13. doi:10.1093/icb/icw020.

Dillon, M.E. *et al.* (2016) 'Life in the frequency domain: The biological impacts of changes in climate variability at multiple time scales', *Integrative and Comparative Biology*, 56(1), pp. 14–30. doi:10.1093/icb/icw024.

Dutta, A. and Zisserman, A. (2019) 'The VIA annotation software for images, audio and video', *Proceedings of the 27th ACM International Conference on Multimedia* [Preprint]. doi:10.1145/3343031.3350535.

Easterling, D.R. *et al.* (2000b) 'Climate extremes: Observations, modeling, and impacts', *Science*, 289(5487), pp. 2068–2074. doi:10.1126/science.289.5487.2068.

Edelstein, A. *et al.* (2010) 'Computer control of microscopes using μManager', *Current Protocols in Molecular Biology*, 92(1). doi:10.1002/0471142727.mb1420s92.

Feichtenhofer, C. *et al.* (2019) 'SLOWFAST networks for video recognition', *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* [Preprint]. doi:10.1109/iccv.2019.00630.

Fink, M. *et al.* (2009) 'A new method for detection and quantification of heartbeat parameters in drosophila, zebrafish, and embryonic Mouse Hearts', *BioTechniques*, 46(2), pp. 101–113. doi:10.2144/000113078.

Fodor, I. *et al.* (2020) 'The unlimited potential of the great pond snail, Lymnaea stagnalis', *eLife*, 9. doi:10.7554/elife.56962.

Foray, V., Desouhant, E. and Gibert, P. (2013) 'The impact of thermal fluctuations on reaction norms in specialist and generalist parasitic wasps', *Functional Ecology*, 28(2), pp. 411–423. doi:10.1111/1365-2435.12171.

Foster, N.L., Lukowiak, K. and Henry, T.B. (2015) 'Time-related expression profiles for heat shock protein gene transcripts (*HSP40, hsp70*) in the central nervous system of*lymnaea stagnalis*exposed to thermal stress', *Communicative & Integrative Biology*, 8(3). doi:10.1080/19420889.2015.1040954.

Garstang, W. (1922) 'The theory of recapitulation: A critical re-statement of the biogenetic law.', *Journal of the Linnean Society of London, Zoology*, 35(232), pp. 81–101. doi:10.1111/j.1096-3642.1922.tb00464.x.

Genin, A. *et al.* (2020) 'Rapid onsets of warming events trigger mass mortality of coral reef fish', *Proceedings of the National Academy of Sciences*, 117(41), pp. 25378–25385. doi:10.1073/pnas.2009748117.

Georges, A. *et al.* (2005) 'Modelling development of reptile embryos under fluctuating temperature regimes', *Physiological and Biochemical Zoology*, 78(1), pp. 18–30. doi:10.1086/425200.

Ghaedi, B. and Andrew, N.R. (2016) 'The physiological consequences of varied heat exposure events in adult*myzus persicae*: A single prolonged exposure compared to repeated shorter exposures', *PeerJ*, 4. doi:10.7717/peerj.2290.

Gierten, J. *et al.* (2020b) 'Automated high-throughput heartbeat quantification in medaka and zebrafish embryos under physiological conditions', *Scientific Reports*, 10(1). doi:10.1038/s41598-020-58563-w.

Gould, S.J. (1977) *Ontogeny and phylogeny Stephen Jay gould*. Cambridge, Mass. u.a: The Belknap Pr. of Harvard Univ. Pr.

Grinberg, M. (2018) *Flask web development: Developing web applications with python*. Sebastopol, CA: O'Reilly.

Haeckel, E. (1866). Generelle Morphologie der Organism Allgemein Grundzügeder Organischen Formen-Wissenschaft Mechanisch Begründet durch die von Charles Darwin Reformite Descendenz-Theorie (2 vols.). Berlin: George Reimer.

Hagstrum, D.W. and Milliken, G.A. (1991) 'Modeling differences in insect developmental times between constant and fluctuating temperatures', *Annals of the Entomological Society of America*, 84(4), pp. 369–379. doi:10.1093/aesa/84.4.369.

Hall, J.M. and Warner, D.A. (2019) 'Thermal tolerance in the urban heat island: Thermal sensitivity varies ontogenetically and differs between embryos of two sympatric ectotherms', *Journal of Experimental Biology* [Preprint]. doi:10.1242/jeb.210708.

Hallou, A. *et al.* (2021) 'Deep Learning for bioimage analysis in developmental biology', *Development*, 148(18). doi:10.1242/dev.199616.

He, K. *et al.* (2016) 'Deep residual learning for image recognition', *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [Preprint]. doi:10.1109/cvpr.2016.90.

Headlee, T.J. (1914) 'Some data on the effect of temperature and moisture on the rate of insect metabolism', *Journal of Economic Entomology*, 7(6), pp. 413–417. doi:10.1093/jee/7.6.413.

Hillebrand, S. *et al.* (2013) 'Heart rate variability and first cardiovascular event in populations without known cardiovascular disease: Meta-analysis and dose–response meta-regression', *EP Europace*, 15(5), pp. 742–749. doi:10.1093/europace/eus341.

Hoage, T., Ding, Y. and Xu, X. (2011) 'Quantifying cardiac functions in embryonic and adult zebrafish', *Methods in Molecular Biology*, pp. 11–20. doi:10.1007/978-1-61779-523-7_2.

Hollandi, R. *et al.* (2020) 'Annotatorj: An imagej plugin to ease hand annotation of cellular compartments', *Molecular Biology of the Cell*, 31(20), pp. 2179–2186. doi:10.1091/mbc.e20-02-0156.

Houle, D., Govindaraju, D.R. and Omholt, S. (2010) 'Phenomics: The next challenge', *Nature Reviews Genetics*, 11(12), pp. 855–866. doi:10.1038/nrg2897.

Howe, R.W. (1967) 'Temperature effects on embryonic development in insects', *Annual Review of Entomology*, 12(1), pp. 15–42. doi:10.1146/annurev.en.12.010167.000311.

Hubendick, B. (1951) *Recent Lymnaeidae: Their variation, morphology, taxonomy, nomenclature, and Distribution*. Stockholm: Almqvist & Wiksell.

Huey, R.B. and Kearney, M.R. (2020) 'Dynamics of death by heat', *Science*, 369(6508), pp. 1163–1163. doi:10.1126/science.abe0320.

Huey, R.B. and Kingsolver, J.G. (1989) 'Evolution of thermal sensitivity of ectotherm performance', *Trends in Ecology & Evolution*, 4(5), pp. 131–135. doi:10.1016/0169-5347(89)90211-5.

Huey, R.B. and Stevenson, R.D. (1979) 'Integrating thermal physiology and ecology of ectotherms: A discussion of approaches', *American Zoologist*, 19(1), pp. 357–366. doi:10.1093/icb/19.1.357.

Huey, R.B. *et al.* (1999) 'Testing the adaptive significance of acclimation: A strong inference approach', *American Zoologist*, 39(2), pp. 323–336. doi:10.1093/icb/39.2.323.

Huey, R.B. *et al.* (2012) 'Predicting organismal vulnerability to climate warming: Roles of behaviour, physiology and adaptation', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1596), pp. 1665–1679. doi:10.1098/rstb.2012.0005.

Huey, R.B., Hertz, P.E. and Sinervo, B. (2003) 'Behavioral drive versus behavioral inertia in evolution: A null model approach', *The American Naturalist*, 161(3), pp. 357–366. doi:10.1086/346135.

Humpesch, U.H. (1982) 'Effect of fluctuating temperature on the duration of embryonic development in two ecdyonurus spp. and Rhithrogena CF. Hybrida (ephemeroptera) from Austrian streams', *Oecologia*, 55(3), pp. 285–288. doi:10.1007/bf00376913.

Ibbini, Z. *et al.* (2022) 'HeartCV: A tool for transferrable, automated measurement of heart rate and heart rate variability in transparent animals', *Journal of Experimental Biology*, 225(19). doi:10.1242/jeb.244729.

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 [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, pp. 35-115.

Ishaq, O., Sadanandan, S.K. and Wählby, C. (2017) 'Deep fish', *SLAS Discovery*, 22(1), pp. 102–107. doi:10.1177/1087057116667894.

Jensen, J.L. (1906) 'Sur les fonctions convexes et les inégalités entre les valeurs moyennes', *Acta Mathematica*, 30(0), pp. 175–193. doi:10.1007/bf02418571.

Johansson, M.P. and Laurila, A. (2017) 'Maximum thermal tolerance trades off with chronic tolerance of high temperature in contrasting thermal populations of *radix balthica*', *Ecology and Evolution*, 7(9), pp. 3149–3156. doi:10.1002/ece3.2923.

Jones, R.A., Renshaw, M.J. and Barry, D.J. (2023) 'Automated staging of zebrafish embryos with deep learning', *Life Science Alliance*, 7(1). doi:10.26508/lsa.202302351.

Jørgensen, L.B. *et al.* (2021) 'Dramatic changes in mitochondrial substrate use at critically high temperatures: A comparative study using *drosophila*', *Journal of Experimental Biology*, 224(6). doi:10.1242/jeb.240960.

Jørgensen, L.B., Malte, H. and Overgaard, J. (2019) 'How to assess *drosophila* heat tolerance: Unifying static and dynamic tolerance assays to predict heat distribution limits', *Functional Ecology*, 33(4), pp. 629–642. doi:10.1111/1365-2435.13279.

Kaufmann, O. (1932) 'Einige Bemerkungen Ber Den Einfluss von Temperaturschwankungen auf die entwicklungsdauer und streuung bei insekten und seine graphische Darstellung durch Kettenlinie und hyperbel', *Zeitschrift fr Morphologie und kologie der Tiere*, 25(2–3), pp. 353–361. doi:10.1007/bf00446713.

Kearney, M., Shine, R. and Porter, W.P. (2009) 'The potential for behavioral thermoregulation to buffer "cold-blooded" animals against climate warming', *Proceedings of the National Academy of Sciences*, 106(10), pp. 3835–3840. doi:10.1073/pnas.0808913106.

Keen, R. (1979) 'Effects of fluctuating temperature on duration on egg development of Chydorus Sphaericus (Cladocera, Crustacea)', *Journal of Thermal Biology*, 4(1), pp. 5–8. doi:10.1016/0306-4565(79)90038-x.

Keyser, C.A. *et al.* (2014) 'Heat-induced post-stress growth delay: A biological trait of many metarhizium isolates reducing biocontrol efficacy?', *Journal of Invertebrate Pathology*, 120, pp. 67–73. doi:10.1016/j.jip.2014.05.008.

Khelifa, R. *et al.* (2019) 'Usefulness and limitations of thermal performance curves in predicting ectotherm development under climatic variability', *Journal of Animal Ecology*, 88(12), pp. 1901–1912. doi:10.1111/1365-2656.13077.

Khosravi, P. *et al.* (2019) 'Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization', *npj Digital Medicine*, 2(1). doi:10.1038/s41746-019-0096-y.

Kingsolver, J.G. (2004) 'Plasticity of size and growth in fluctuating thermal environments: Comparing reaction norms and performance curves', *Integrative and Comparative Biology*, 44(6), pp. 450–460. doi:10.1093/icb/44.6.450.

Kingsolver, J. and Huey, R., 2008. Size, temperature, and fitness: three rules. Evolutionary Ecology Research, 10(2), pp.251-268.

Kingsolver, J.G. and Woods, H.A. (2016) 'Beyond thermal performance curves: Modeling time-dependent effects of thermal stress on ectotherm growth rates', *The American Naturalist*, 187(3), pp. 283–294. doi:10.1086/684786.

Kingsolver, J.G. *et al.* (2020) 'Growth, stress, and acclimation responses to fluctuating temperatures in field and domesticated populations of *manduca sexta*', *Ecology and Evolution*, 10(24), pp. 13980–13989. doi:10.1002/ece3.6991.

Kingsolver, J.G., Diamond, S.E. and Buckley, L.B. (2013) 'Heat stress and the fitness consequences of climate change for terrestrial ectotherms', *Functional Ecology*, 27(6), pp. 1415–1423. doi:10.1111/1365-2435.12145.

Kingsolver, J.G., Higgins, J.K. and Augustine, K.E. (2015) 'Fluctuating temperatures and ectotherm growth: Distinguishing non-linear and time-dependent effects', *Journal of Experimental Biology* [Preprint]. doi:10.1242/jeb.120733.

Klockmann, M., Günter, F. and Fischer, K. (2016) 'Heat resistance throughout ontogeny: Body size constrains thermal tolerance', *Global Change Biology*, 23(2), pp. 686–696. doi:10.1111/gcb.13407.

Kobayashi, S. *et al.* (2017) 'The effects of nest incubation temperature on embryos and hatchlings of the loggerhead sea turtle: Implications of sex difference for survival rates during early life stages', *Journal of Experimental Marine Biology and Ecology*, 486, pp. 274–281. doi:10.1016/j.jembe.2016.10.020.

Komoroske, L.M. *et al.* (2014) 'Ontogeny influences sensitivity to climate change stressors in an endangered fish', *Conservation Physiology*, 2(1). doi:10.1093/conphys/cou008.

Kondratyuk, D. *et al.* (2021) 'Movinets: Mobile Video Networks for Efficient Video Recognition', *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* [Preprint]. doi:10.1109/cvpr46437.2021.01576.

Kregel, K.C. (2002) 'Invited review: Heat shock proteins: Modifying factors in physiological stress responses and acquired thermotolerance', *Journal of Applied Physiology*, 92(5), pp. 2177–2186. doi:10.1152/japplphysiol.01267.2001.

Krenek, S., Berendonk, T.U. and Petzoldt, T. (2011) 'Thermal performance curves of paramecium caudatum: A model selection approach', *European Journal of Protistology*, 47(2), pp. 124–137. doi:10.1016/j.ejop.2010.12.001.

Krishna, S., Chatti, K. and Galigekere, R.R. (2018) 'Automatic and robust estimation of heart rate in zebrafish larvae', *IEEE Transactions on Automation Science and Engineering*, 15(3), pp. 1041–1052. doi:10.1109/tase.2017.2705240.

Kurnia, K.A. *et al.* (2021) 'Measurement of multiple cardiac performance endpoints in daphnia and zebrafish by Kymograph', *Inventions*, 6(1), p. 8. doi:10.3390/inventions6010008.

Kuroda, R. and Abe, M. (2020) 'The pond snail Lymnaea stagnalis', *EvoDevo*, 11(1). doi:10.1186/s13227-020-00169-4.

Lagerspetz, K.Y.H. (2006) 'What is thermal acclimation?', *Journal of Thermal Biology*, 31(4), pp. 332–336. doi:10.1016/j.jtherbio.2006.01.003.

Laine, R.F. *et al.* (2021) 'Avoiding a replication crisis in deep-learning-based bioimage analysis', *Nature Methods*, 18(10), pp. 1136–1144. doi:10.1038/s41592-021-01284-3.

LeCun, Y., Bengio, Y. and Hinton, G. (2015) 'Deep learning', *Nature*, 521(7553), pp. 436–444. doi:10.1038/nature14539.

Lee, R.E. (1989) 'Insect cold-hardiness: To freeze or not to freeze', *BioScience*, 39(5), pp. 308–313. doi:10.2307/1311113.

Lerin, J. (2004) 'Modeling embryonic development in *sitona lineatus* (Coleoptera: Curculionidae) in fluctuating temperatures', *Environmental Entomology*, 33(2), pp. 107–112. doi:10.1603/0046-225x-33.2.107.

Levy, O. *et al.* (2015) 'Resolving the life cycle alters expected impacts of climate change', *Proceedings of the Royal Society B: Biological Sciences*, 282(1813), p. 20150837. doi:10.1098/rspb.2015.0837.

Li, X. *et al.* (2019) 'Zebrafish behavioral phenomics applied for phenotyping aquatic neurotoxicity induced by lead contaminants of environmentally relevant level', *Chemosphere*, 224, pp. 445–454. doi:10.1016/j.chemosphere.2019.02.174.

Liu, W. *et al.* (2022) 'Embryonic exposure to selenium nanoparticles delays growth and hatching in the freshwater snail Lymnaea stagnalis', *Chemosphere*, 307, p. 136147. doi:10.1016/j.chemosphere.2022.136147.

Liu, Z. *et al.* (2019) 'Multi-task deep learning with dynamic programming for embryo early development stage classification from time-lapse videos', *IEEE Access*, 7, pp. 122153–122163. doi:10.1109/access.2019.2937765.

Louis, C.M. *et al.* (2021) 'Review of computer vision application in in vitro fertilization: The application of Deep Learning-based Computer Vision Technology in the world of IVF', *Journal of Assisted Reproduction and Genetics*, 38(7), pp. 1627–1639. doi:10.1007/s10815-021-02123-2.

Love, A.C. (2010) 'Idealization in evolutionary developmental investigation: A tension between phenotypic plasticity and normal stages', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1540), pp. 679–690. doi:10.1098/rstb.2009.0262.

Ludwig, D. and Cable, R.M. (1933) 'The effect of alternating temperatures on the pupal development of drosophila melanogaster meigen', *Physiological Zoology*, 6(4), pp. 493–508. doi:10.1086/physzool.6.4.30151203.

Machikhin, A.S. *et al.* (2020) 'Blood vessel imaging at pre-larval stages of zebrafish embryonic development', *Diagnostics*, 10(11), p. 886. doi:10.3390/diagnostics10110886.

Madeira, D. *et al.* (2015) 'Physiological, cellular and biochemical thermal stress response of intertidal shrimps with different vertical distributions: Palaemon elegans and palaemon serratus', *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 183, pp. 107–115. doi:10.1016/j.cbpa.2014.12.039.

Malik, M. *et al.* (1996) 'Heart rate variability: Standards of measurement, physiological interpretation, and clinical use', *European Heart Journal*, 17(3), pp. 354–381. doi:10.1093/oxfordjournals.eurheartj.a014868.

Marshall, K.E. and Sinclair, B.J. (2014) 'The relative importance of number, duration and intensity of cold stress events in determining survival and energetics of an overwintering insect', *Functional Ecology*, 29(3), pp. 357–366. doi:10.1111/1365-2435.12328.

Marshall, K.E. *et al.* (2021) 'Whole-organism responses to constant temperatures do not predict responses to variable temperatures in the ecosystem engineer *mytilus trossulus*', *Proceedings of the Royal Society B: Biological Sciences*, 288(1947). doi:10.1098/rspb.2020.2968.

McCalla, K.A. *et al.* (2019) 'The influence of temperature variation on life history parameters and thermal performance curves of Tamarixia radiata (hymenoptera: Eulophidae), a parasitoid of the Asian citrus psyllid (Hemiptera: Liviidae)', *Journal of Economic Entomology*, 112(4), pp. 1560–1574. doi:10.1093/jee/toz067.

McCoy, J.C. *et al.* (2023) 'Phenomics as an approach to comparative developmental physiology', *Frontiers in Physiology*, 14. doi:10.3389/fphys.2023.1229500.

McInnes, L. *et al.* (2018) 'UMAP: Uniform manifold approximation and projection', *Journal of Open Source Software*, 3(29), p. 861. doi:10.21105/joss.00861.

McKinney, M.L. and McNamara, K. (1991) *Heterochrony: The Evolution of Ontogeny*. New York: Plenum Press.

McNamara, K.J. and McKinney, M.L. (2005) 'Heterochrony, disparity, and macroevolution', *Paleobiology*, 31(sp5), pp. 17–26. doi:10.1666/0094-8373(2005)031[0017:hdam]2.0.co;2.

Meehl, G.A. and Tebaldi, C. (2004) 'More intense, more frequent, and longer lasting heat waves in the 21st century', *Science*, 305(5686), pp. 994–997. doi:10.1126/science.1098704.

Meijering, E. (2020) 'A bird's-eye view of deep learning in Bioimage analysis', *Computational and Structural Biotechnology Journal*, 18, pp. 2312–2325. doi:10.1016/j.csbj.2020.08.003.

Meijering, E. *et al.* (2016) 'Imagining the future of bioimage analysis', *Nature Biotechnology*, 34(12), pp. 1250–1255. doi:10.1038/nbt.3722.

Melendez, C.L. and Mueller, C.A. (2021) 'Effect of increased embryonic temperature during developmental windows on survival, morphology and oxygen consumption of rainbow

trout (Oncorhynchus mykiss)', *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 252, p. 110834. doi:10.1016/j.cbpa.2020.110834.

Meshcheryakov, V.N. (1990) 'The common pond snail Lymnaea stagnalis', *Animal Species for Developmental Studies*, pp. 69–132. doi:10.1007/978-1-4613-0503-3_5.

Messenger, P.S. and Flitters, N.E. (1959) 'Effect of variable temperature environments on egg development of three species of fruit flies1', *Annals of the Entomological Society of America*, 52(2), pp. 191–204. doi:10.1093/aesa/52.2.191.

Morash, A.J. *et al.* (2018a) 'The importance of incorporating natural thermal variation when evaluating physiological performance in wild species', *Journal of Experimental Biology*, 221(14). doi:10.1242/jeb.164673.

Morash, A.J. *et al.* (2018b) 'The importance of incorporating natural thermal variation when evaluating physiological performance in wild species', *Journal of Experimental Biology*, 221(14). doi:10.1242/jeb.164673.

Morash, A.J. *et al.* (2021) 'The physiological ups and downs of thermal variability in temperate freshwater ecosystems', *Journal of Fish Biology*, 98(6), pp. 1524–1535. doi:10.1111/jfb.14655.

Mousavi, S.E. and Patil, J.G. (2020) 'Light-cardiogram, a simple technique for heart rate determination in adult zebrafish, Danio rerio', *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 246, p. 110705. doi:10.1016/j.cbpa.2020.110705.

Mueller, C.A. *et al.* (2015) 'Embryonic critical windows: Changes in incubation temperature alter survival, hatchling phenotype, and cost of development in Lake Whitefish (Coregonus

clupeaformis)', *Journal of Comparative Physiology B*, 185(3), pp. 315–331. doi:10.1007/s00360-015-0886-8.

Myers, G. (2012) 'Why Bioimage Informatics Matters', *Nature Methods*, 9(7), pp. 659–660. doi:10.1038/nmeth.2024.

Nachman, G. and Skovgård, H. (2020) 'Modeling the influence of ambient temperature on the interactions between the stable fly (Diptera: Muscidae) and its natural enemy Spalangia Cameroni (hymenoptera: Pteromalidae) to assess consequences of climate change', *Environmental Entomology*, 49(2), pp. 342–354. doi:10.1093/ee/nvaa008.

Nakadera, Y., Blom, C. and Koene, J.M. (2014) 'Duration of sperm storage in the simultaneous hermaphrodite Lymnaea stagnalis', *Journal of Molluscan Studies*, 80(1), pp. 1–7. doi:10.1093/mollus/eyt049.

Niehaus, A.C. *et al.* (2012) 'Predicting the physiological performance of ectotherms in fluctuating thermal environments', *Journal of Experimental Biology*, 215(4), pp. 694–701. doi:10.1242/jeb.058032.

Nogare, D.D. *et al.* (2023) 'Using AI in bioimage analysis to elevate the rate of scientific discovery as a community', *Nature Methods*, 20(7), pp. 973–975. doi:10.1038/s41592-023-01929-5.

Noldus, L.P., Spink, A.J. and Tegelenbosch, R.A. (2001) 'EthoVision: A versatile video tracking system for automation of behavioral experiments', *Behavior Research Methods, Instruments, & Computers*, 33(3), pp. 398–414. doi:10.3758/bf03195394.

Norman, G.J. *et al.* (2012) 'Heart rate variability predicts cell death and inflammatory responses to global cerebral ischemia', *Frontiers in Physiology*, 3. doi:10.3389/fphys.2012.00131.

Olsson, M. and Uller, T. (2002) 'Developmental stability and genetic architecture: A comparison within and across thermal regimes in tadpoles', *Journal of Evolutionary Biology*, 15(4), pp. 625–633. doi:10.1046/j.1420-9101.2002.00417.x.

Otsu, N. (1979) 'A threshold selection method from gray-level histograms', *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), pp. 62–66. doi:10.1109/tsmc.1979.4310076.

O'Mahony, N. *et al.* (2019) 'Deep learning vs. Traditional Computer Vision', *Advances in Intelligent Systems and Computing*, pp. 128–144. doi:10.1007/978-3-030-17795-9_10.

Padfield, D., O'Sullivan, H. and Pawar, S. (2021) '*rtpc* and *nls.multstart*: A new pipeline to fit thermal performance curves in r', *Methods in Ecology and Evolution*, 12(6), pp. 1138–1143. doi:10.1111/2041-210x.13585.

Pandori, L.L. and Sorte, C.J. (2018) 'The weakest link: Sensitivity to climate extremes across life stages of marine invertebrates', *Oikos*, 128(5), pp. 621–629. doi:10.1111/oik.05886.

Parker, J.R., 1930. Some effects of temperature and moisture upon *Melanoplus mexicanus Saussure* and *Camnula pellucida Scudder* (Orthoptera). Montana Agric. Expt. Sta. Bull. 223. 132 pp.

Paternostro, G. *et al.* (2001) 'Age-associated cardiac dysfunction in *drosophila melanogaster*', *Circulation Research*, 88(10), pp. 1053–1058. doi:10.1161/hh1001.090857.

Paul, R.J. *et al.* (1997) 'Circulation and respiratory control in millimetre-sized animals (Daphnia Magna , Folsomia Candida) studied by optical methods', *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology*, 167(6), pp. 399–408. doi:10.1007/s003600050089.

Peng, H. (2008) 'Bioimage Informatics: A new area of engineering biology', *Bioinformatics*, 24(17), pp. 1827–1836. doi:10.1093/bioinformatics/btn346.

Pond, A.J. *et al.* (2021) 'A deep learning approach for staging embryonic tissue isolates with small data', *PLOS ONE*, 16(1). doi:10.1371/journal.pone.0244151.

Pottier, P. *et al.* (2022) 'Developmental plasticity in thermal tolerance: Ontogenetic variation, persistence, and Future Directions', *Ecology Letters*, 25(10), pp. 2245–2268. doi:10.1111/ele.14083.

Pörtner, H.O. and Farrell, A.P., 2008. Physiology and climate change. Science, 322(5902), pp.690-692.

Pradhan, S., 1945. Insect population studies. II. Rate of insect development under variable temperature of the field. In Proc. Nat. Inst. Sci. India (Vol. 11, No. 2, pp. 74-80).

Puybareau, E., Talbot, H. and Leonard, M. (2015) 'Automated heart rate estimation in fish embryo', *2015 International Conference on Image Processing Theory, Tools and Applications (IPTA)* [Preprint]. doi:10.1109/ipta.2015.7367171.

Pylatiuk, C. *et al.* (2014) 'Automatic zebrafish heartbeat detection and analysis for zebrafish embryos', *Zebrafish*, 11(4), pp. 379–383. doi:10.1089/zeb.2014.1002.

R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Radchuk, V., Turlure, C. and Schtickzelle, N. (2012) 'Each life stage matters: The importance of assessing the response to climate change over the complete life cycle in butterflies', *Journal of Animal Ecology*, 82(1), pp. 275–285. doi:10.1111/j.1365-2656.2012.02029.x.

Rajendra Acharya, U. *et al.* (2006) 'Heart rate variability: A Review', *Medical & Biological Engineering & Computing*, 44(12), pp. 1031–1051. doi:10.1007/s11517-006-0119-0.

Rezende, E.L. and Bozinovic, F. (2019) 'Thermal performance across levels of biological organization', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1778), p. 20180549. doi:10.1098/rstb.2018.0549.

Rezende, E.L., Castañeda, L.E. and Santos, M. (2014b) 'Tolerance landscapes in thermal ecology', *Functional Ecology*, 28(4), pp. 799–809. doi:10.1111/1365-2435.12268.

Richardson, M.K. and Keuck, G. (2022) 'The revolutionary developmental biology of wilhelm his, sr.', *Biological Reviews*, 97(3), pp. 1131–1160. doi:10.1111/brv.12834.

Roitberg, B.D. and Mangel, M. (2016) 'Cold snaps, heatwaves, and arthropod growth', *Ecological Entomology*, 41(6), pp. 653–659. doi:10.1111/een.12324.

Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing.

Royer, L.A. (2023) 'The future of bioimage analysis: A dialog between mind and machine', *Nature Methods*, 20(7), pp. 951–952. doi:10.1038/s41592-023-01930-y.

Ruel, J.J. and Ayres, M.P. (1999) 'Jensen's inequality predicts effects of environmental variation', *Trends in Ecology & Evolution*, 14(9), pp. 361–366. doi:10.1016/s0169-5347(99)01664-x.

Ruhr, I.M. *et al.* (2019) 'Developmental plasticity of cardiac anoxia-tolerance in juvenile common snapping turtles (*chelydra serpentina*)', *Proceedings of the Royal Society B: Biological Sciences*, 286(1905), p. 20191072. doi:10.1098/rspb.2019.1072.

Rummukainen, M. (2012) 'Changes in climate and weather extremes in the 21st Century', *WIREs Climate Change*, 3(2), pp. 115–129. doi:10.1002/wcc.160.

Rundle, S.D. and Spicer, J.I. (2016) 'Heterokairy: A significant form of developmental plasticity?', *Biology Letters*, 12(9), p. 20160509. doi:10.1098/rsbl.2016.0509.

Ruthsatz, K. *et al.* (2022b) 'Thermal tolerance and acclimation capacity in the European common frog (*rana temporaria*) change throughout ontogeny', *Journal of Experimental Zoology Part A: Ecological and Integrative Physiology*, 337(5), pp. 477–490. doi:10.1002/jez.2582.

Salman, H.E. and Yalcin, H.C. (2020) 'Advanced blood flow assessment in zebrafish via experimental digital particle image velocimetry and computational fluid dynamics modeling', *Micron*, 130, p. 102801. doi:10.1016/j.micron.2019.102801.

Sandehson, D.E. (1910) 'The relation of temperature to the growth of insects', *Journal of Economic Entomology*, 3(2), pp. 113–140. doi:10.1093/jee/3.2.113.

Santoso, F., Farhan, A., *et al.* (2020) 'An overview of methods for cardiac rhythm detection in zebrafish', *Biomedicines*, 8(9), p. 329. doi:10.3390/biomedicines8090329.

Santoso, F., Krylov, V.V., *et al.* (2020) 'Cardiovascular performance measurement in water fleas by utilizing high-speed videography and imagej software and its application for pesticide toxicity assessment', *Animals*, 10(9), p. 1587. doi:10.3390/ani10091587.

Sato, A., Shimeld, S.M. and Bishop, J.D. (2014) 'Symmetrical reproductive compatibility of two species in the Ciona intestinalis (ascidiacea) species complex, a model for Marine Genomics and Developmental Biology', *Zoological Science*, 31(6), p. 369. doi:10.2108/zs130249.

Schaum, C.E. *et al.* (2018) 'Temperature-driven selection on metabolic traits increases the strength of an algal–grazer interaction in naturally warmed streams', *Global Change Biology*, 24(4), pp. 1793–1803. doi:10.1111/gcb.14033.

Schechtman, V.L. *et al.* (1992) 'Dynamic analysis of cardiac R-R intervals in normal infants and in infants who subsequently succumbed to the sudden infant death syndrome', *Pediatric Research*, 31(6), pp. 606–612. doi:10.1203/00006450-199206000-00014.

Schindelin, J. *et al.* (2012) 'Fiji: An open-source platform for biological-image analysis', *Nature Methods*, 9(7), pp. 676–682. doi:10.1038/nmeth.2019.

Scholkmann, F., Boss, J. and Wolf, M. (2012) 'An efficient algorithm for automatic peak detection in noisy periodic and quasi-periodic signals', *Algorithms*, 5(4), pp. 588–603. doi:10.3390/a5040588.

Schroff, F., Kalenichenko, D. and Philbin, J. (2015) 'FaceNet: A unified embedding for face recognition and clustering', *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [Preprint]. doi:10.1109/cvpr.2015.7298682.

Schulte, P.M., Healy, T.M. and Fangue, N.A. (2011) 'Thermal performance curves, phenotypic plasticity, and the time scales of temperature exposure', *Integrative and Comparative Biology*, 51(5), pp. 691–702. doi:10.1093/icb/icr097.

Serafini, L. *et al.* (2011) 'The proteomic response of sea squirts (genus Ciona) to acute heat stress: A global perspective on the thermal stability of proteins', *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 6(3), pp. 322–334. doi:10.1016/j.cbd.2011.07.002.

Seuront, L. *et al.* (2019) 'Decreased thermal tolerance under recurrent heat stress conditions explains summer mass mortality of the Blue Mussel Mytilus edulis', *Scientific Reports*, 9(1). doi:10.1038/s41598-019-53580-w.

Shariff, A. *et al.* (2010) 'Automated Image Analysis for high-content screening and analysis', *SLAS Discovery*, 15(7), pp. 726–734. doi:10.1177/1087057110370894.

Shelford, V. (1927) 'An experimental investigation of the relations of the codling moth to weather and climate', *Illinois Natural History Survey Bulletin*, 16(1–7), pp. 312–440. doi:10.21900/j.inhs.v16.296.

Siddiqui, W.H. and Barlow, C.A. (1972) 'Population growth of drosophila melanogaster (Diptera: Drosophilidae) at constant and alternating temperatures1', *Annals of the Entomological Society of America*, 65(5), pp. 993–1001. doi:10.1093/aesa/65.5.993.

Sinclair, B.J. *et al.* (2016) 'Can we predict ectotherm responses to climate change using thermal performance curves and body temperatures?', *Ecology Letters*, 19(11), pp. 1372–1385. doi:10.1111/ele.12686.

Smale, D.A. *et al.* (2019) 'Marine heatwaves threaten global biodiversity and the provision of Ecosystem Services', *Nature Climate Change*, 9(4), pp. 306–312. doi:10.1038/s41558-019-0412-1.

Smirthwaite, J.J. *et al.* (2007) 'An integrative approach identifies developmental sequence heterochronies in freshwater basommatophoran snails', *Evolution & Development*, 9(2), pp. 122–130. doi:10.1111/j.1525-142x.2007.00143.x.

Smith, L.A. and Lancaster, L.T. (2020) 'Increased duration of extreme thermal events negatively affects cold acclimation ability in a high-latitude, freshwater ectotherm (Ischnura

elegans; Odonata: Coenagrionidae)', *European Journal of Entomology*, 117, pp. 93–100. doi:10.14411/eje.2020.010.

Soule, M. and Yang, S.Y. (1973) 'Genetic variation in side-blotched lizards on islands in the Gulf of California', Evolution, 27(4), p. 593. doi:10.2307/2407193.

Spicer, J. and Rundle, S. (2006) 'Out of place and out of time – towards a more integrated approach to heterochrony', *Animal Biology*, 56(4), pp. 487–502. doi:10.1163/157075606778967810.

Spicer, J.I. (1995) 'Ontogeny of respiratory function in crustaceans exhibiting either direct or indirect development', *Journal of Experimental Zoology*, 272(6), pp. 413–418. doi:10.1002/jez.1402720602.

Spicer, J.I., Rundle, S.D. and Tills, O. (2011) 'Studying the altered timing of physiological events during development: It's about time…or is it?', *Respiratory Physiology & Neurobiology*, 178(1), pp. 3–12. doi:10.1016/j.resp.2011.06.005.

Spomer, W. *et al.* (2012b) 'High-throughput screening of zebrafish embryos using automated heart detection and imaging', *SLAS Technology*, 17(6), pp. 435–442. doi:10.1177/2211068212464223.

Stein, P.K. *et al.* (1994) 'Heart rate variability: A measure of cardiac autonomic tone', *American Heart Journal*, 127(5), pp. 1376–1381. doi:10.1016/0002-8703(94)90059-0.

Storch, D. *et al.* (2011) 'Thermal tolerance of larval stages of the Chilean kelp crab taliepus dentatus ', *Marine Ecology Progress Series*, 429, pp. 157–167. doi:10.3354/meps09059.

Swedlow, J.R., Goldberg, I.G. and Eliceiri, K.W. (2009a) 'Bioimage informatics for Experimental Biology', *Annual Review of Biophysics*, 38(1), pp. 327–346. doi:10.1146/annurev.biophys.050708.133641.

Szeliski, R. (2011) *Computer vision algorithms and applications*. London: Springer London.

Sørensen, J.G., Kristensen, T.N. and Loeschcke, V. (2003) 'The evolutionary and ecological role of heat shock proteins', *Ecology Letters*, 6(11), pp. 1025–1037. doi:10.1046/j.1461-0248.2003.00528.x.

Tanigoshi, L.K. and Logan, J.A. (1979) 'TETRANYCHID development under variable temperature regimes', *Recent Advances in Acarology*, pp. 165–175. doi:10.1016/b978-0-12-592201-2.50028-2.

Tardieu, F. *et al.* (2017) 'Plant phenomics, from sensors to knowledge', *Current Biology*, 27(15). doi:10.1016/j.cub.2017.05.055.

Taylor, P.S. and Shields, E.J. (1990) 'Development of the Armyworm (Lepidoptera: Noctuidae) under fluctuating daily temperature regimes', *Environmental Entomology*, 19(5), pp. 1422–1431. doi:10.1093/ee/19.5.1422.

Terblanche, J.S. *et al.* (2007) 'Critical thermal limits depend on methodological context', *Proceedings of the Royal Society B: Biological Sciences*, 274(1628), pp. 2935–2943. doi:10.1098/rspb.2007.0985.

Thomson, J.A. *et al.* (2014) 'Extreme temperatures, foundation species, and abrupt ecosystem change: An example from an iconic seagrass ecosystem', *Global Change Biology*, 21(4), pp. 1463–1474. doi:10.1111/gcb.12694.

Tills, O. *et al.* (2018) 'A high-throughput and open-source platform for embryo phenomics', *PLOS Biology*, 16(12). doi:10.1371/journal.pbio.3000074.

Tills, O. *et al.* (2021) 'Spectral phenotyping of embryonic development reveals integrative thermodynamic responses', *BMC Bioinformatics*, 22(1). doi:10.1186/s12859-021-04152-1.

Tills, O., Spicer, J. and Rundle, S. (2010) 'Salinity-induced heterokairy in an upper-estuarine population of the Snail Radix Balthica (mollusca: Pulmonata)', *Aquatic Biology*, 9, pp. 95–105. doi:10.3354/ab00231.

Tomanek, L. (2010) 'Variation in the heat shock response and its implication for predicting the effect of global climate change on species' biogeographical distribution ranges and metabolic costs', *Journal of Experimental Biology*, 213(6), pp. 971–979. doi:10.1242/jeb.038034.

Toulany, N. *et al.* (2023) 'Uncovering developmental time and tempo using Deep Learning', *Nature Methods*, 20(12), pp. 2000–2010. doi:10.1038/s41592-023-02083-8.

Tran, D. *et al.* (2018) 'A closer look at spatiotemporal convolutions for action recognition', *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* [Preprint]. doi:10.1109/cvpr.2018.00675.

Truebano, M. *et al.* (2018b) 'Thermal strategies vary with life history stage', *Journal of Experimental Biology*, 221(8). doi:10.1242/jeb.171629.

Tsuji, H. *et al.* (1994) 'Reduced heart rate variability and mortality risk in an elderly cohort. The Framingham Heart Study.', *Circulation*, 90(2), pp. 878–883. doi:10.1161/01.cir.90.2.878.

Turriago, J.L., Parra, C.A. and Bernal, M.H. (2015) 'Upper thermal tolerance in anuran embryos and tadpoles at constant and variable peak temperatures', *Canadian Journal of Zoology*, 93(4), pp. 267–272. doi:10.1139/cjz-2014-0254.

Uvarov, B.P. (1931) *Insects and climate*. London: Entomological Society of London.

Vajedsamiei, J. *et al.* (2021) 'Cyclic thermal fluctuations can be burden or relief for an ectotherm depending on fluctuations' average and amplitude', *Functional Ecology*, 35(11), pp. 2483–2496. doi:10.1111/1365-2435.13889.

Valladares, F. *et al.* (2014) 'The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change', *Ecology Letters*, 17(11), pp. 1351–1364. doi:10.1111/ele.12348.

Vasseur, D.A. *et al.* (2014) 'Increased temperature variation poses a greater risk to species than climate warming', *Proceedings of the Royal Society B: Biological Sciences*, 281(1779), p. 20132612. doi:10.1098/rspb.2013.2612.

Virtanen, P. *et al.* (2020) 'SciPy 1.0: Fundamental algorithms for scientific computing in python', *Nature Methods*, 17(3), pp. 261–272. doi:10.1038/s41592-019-0686-2.

von Schmalensee, L. *et al.* (2021) 'Thermal performance under constant temperatures can accurately predict insect development times across naturally variable microclimates', *Ecology Letters*, 24(8), pp. 1633–1645. doi:10.1111/ele.13779.

Wagner, T.L., Wu, H.-I., Sharpe, P.J., Schoolfield, R.M., *et al.* (1984) 'Modeling insect development rates: A literature review and application of a biophysical model', *Annals of the Entomological Society of America*, 77(2), pp. 208–220. doi:10.1093/aesa/77.2.208.

Walter, T. *et al.* (2010b) 'Visualization of image data from cells to organisms', *Nature Methods*, 7(S3). doi:10.1038/nmeth.1431.

Wolf, E. (1932) 'Pulsation frequency of the advisceral and ABVISCERAL heart beats of Ciona intestinalis in relation to temperature', *Journal of General Physiology*, 16(1), pp. 89–98. doi:10.1085/jgp.16.1.89.

Wong, C.C. *et al.* (2010) 'Non-invasive imaging of human embryos before embryonic genome activation predicts development to the blastocyst stage', *Nature Biotechnology*, 28(10), pp. 1115–1121. doi:10.1038/nbt.1686.

Woods, H.A., Dillon, M.E. and Pincebourde, S. (2015) 'The roles of microclimatic diversity and of behavior in mediating the responses of ectotherms to climate change', *Journal of Thermal Biology*, 54, pp. 86–97. doi:10.1016/j.jtherbio.2014.10.002.

Worner, S.P. (1992) 'Performance of phenological models under variable temperature regimes: Consequences of the Kaufmann or rate summation effect', *Environmental Entomology*, 21(4), pp. 689–699. doi:10.1093/ee/21.4.689.

Xiaobo Zhou and Wong, S.T.C. (2006) 'Informatics challenges of high-throughput microscopy', *IEEE Signal Processing Magazine*, 23(3), pp. 63–72. doi:10.1109/msp.2006.1628879.

Yang, W. *et al.* (2020) 'Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives', *Molecular Plant*, 13(2), pp. 187–214. doi:10.1016/j.molp.2020.01.008.

Yozzo, K.L. *et al.* (2013) 'High-content screening assay for identification of chemicals impacting cardiovascular function in zebrafish embryos', *Environmental Science & Technology*, 47(19), pp. 11302–11310. doi:10.1021/es403360y.

Zabihihesari, A. *et al.* (2021) 'Open access tool and microfluidic devices for phenotypic quantification of heart function of intact fruit fly and zebrafish larvae', *Computers in Biology and Medicine*, 132, p. 104314. doi:10.1016/j.compbiomed.2021.104314.

Zena, L.A. *et al.* (2021) 'It takes time to heal a broken heart: Ventricular plasticity improves heart performance after myocardial infarction in rainbow trout, *oncorhynchus mykiss*', *Journal of Experimental Biology*, 224(23). doi:10.1242/jeb.243578.

Zhang, P. *et al.* (2016) 'The effect of temperature on herbivory by the omnivorous ectotherm Snail Lymnaea stagnalis', *Hydrobiologia*, 812(1), pp. 147–155. doi:10.1007/s10750-016-2891-7.

'*a preliminary study of the thermal requirements of desert reptiles. bulletin of the American Museum of Natural History, volume 83: Article 5*. Raymond Bridgman Cowles , Charles Mitchill Bogert' (1945) *The Quarterly Review of Biology*, 20(2), pp. 170–170. doi:10.1086/394795.