Collection and Analysis of Disease Vector Thermal Responses

Abstract

In the study of vector-borne diseases, vectors receive limited attention, however by applying ecology based mechanistic models we can better understand and predict how vector dynamics respond to temperature, which is especially important when faced with climate change. There are three main objectives to this project: the collection and homogenisation of separate datasets, the analysis of the structure and quality of the unified dataset and an analysis of the variation in thermal response between traits, genus and species. Overall the collected data has an uneven distribution amongst categories, and the data quality of many data series is poor. Analysis found that there is a significant difference between the trait thermal responses. At the genus level this dataset's *Aedes* thermal response of and that of other studies' *Anopheles* response is very similar, so there is not likely to be a significant difference between the two. At the species level there is no significant difference between the thermal responses of *Aedes aegypti* and *Aedes albopictus*, as well as that of *Glossina morsitans* and *Glossina pallidipes.*

Introduction

Vector-borne diseases are widespread, malaria alone causes an estimated 300 million cases a year and dengue another 50-100 million cases. Alongside diseases like sleeping sickness, yellow fever, West Nile fever and Lyme disease, vector-borne diseases account for 17% of the total global disease burden (Tabachnick, 2010). In addition vector-borne diseases are not limited to humans, but are also relevant to agriculture, due to plant and animal pathogens, and to conservation through diseases like avian malaria.

Within the field of epidemiology many of the vector-borne disease studies have been focused on the host, pathogen, and host-pathogen interactions. Often neglected is the fact that vector behaviour and population dynamics play a key role in the epidemiology of the disease. Vector-borne diseases for the most part rely on arthropod vectors like mosquitos, flies and ticks for transmission (Githeko et al., 2000).

The Metabolic theory of ecology establishes a general framework to study how temperature governs a number of individual and population-level traits including density and growth rates through its effects on metabolic rates(Brown et al., 2004; Gillooly et al., 2001). As arthropods are endothermic, environmental temperatures are a good approximate of internal temperature. The presence or absence of the vector acts as a constraint to transmission, and vector population dynamics contribute to transmission intensity. Therefore, many key vector behaviours (e.g. biting rate) and life-history (e.g. fecundity) traits are temperature dependant. As a result, temperature directly affects vector equilibrium population sizes as well as dynamics over time (Beck-Johnson et al., 2013). This is important to consider, because vector abundance strongly constraints disease transmission rates (Mordecai et al., 2013). In particular the vector-pathogen relationship is affected by traits like incubation rate within the vector and host-vector transmission dynamics through bite rate and vector competence (Mordecai et al., 2013).

A thorough meta-analysis of how vectors traits are affected by temperature can lead to a better understanding of the relationship between temperature and vectors. This project will aim to collect the separate data sets that have been created, then after sorting and standardisation of the data, combining them into a single homogenised dataset. The dataset can then be expanded by extracting

new data from papers, possibly including data on plant and animal disease vectors. The synchronised dataset will then be analysed to determine the type and quality of the data present, thermal response models will also be fitted to the data. This is followed by an analysis of the variation in thermal response between traits, genus and species.

In the short term temperatures shift due to seasonality (Mordecai et al., 2013), or from irregular events like heat waves which can result in epidemic outbreaks (Githeko et al., 2000). In the long term shifts in temperature due to climate change is a concern, because most vector-borne diseases are currently limited to tropical and sub-tropical regions, but climate change is expected to alter many vectors' fundamental niche location and size, as well as their population dynamics (Campbell et al., 2015; Githeko et al., 2000). Hence it is important to understand how temperature affects vectors and predict the short and long term changes, which is where ecology based mechanistic models can be utilised. They can be used to predict how temperature affects various aspects of vector biology, e.g. population dynamics, behaviour, transmission rates, and ultimately the basic reproductive number (R_0) of the disease (Mordecai et al., 2013). Many of these aspects are important to disease control, and the information derived from models can be used to help plan future disease control strategies, increasing effectiveness of interventions and reducing costs.

In many cases when creating new models (Beck-Johnson et al., 2013; Mordecai et al., 2013; Martin & Huey, 2008) or comparing models (Lunde, Bayoh & Lindtjorn, 2013) datasets are required for both model parametrisation and model validation. However vector thermal responses can be difficult to acquire even for well-studied vectors, and often involves searching and extracting from primary literature or requesting original data, all of which would need to be coerced into a common format for usage. This process is time-consuming and tedious, by collecting the various datasets into the single collected dataset, this project will also provide a larger pool of commonly formatted data that can be utilised in future research. The distribution of the data within the different categories can an also be used to identify gaps in the available data, which can inform data acquisition planning by revealing areas that would benefit most from future research. Expanding the available data and evaluating the quality can help reduce the uncertainty of the data, which can have a significant effect on model predictions (Johnson et al., 2015).

An analysis of the generality of thermal responses will help determine the applicability of a data series for research, and validate/invalidate assumptions of response similarity. Due to the lack of available data some studies have to substitute data from a different genus or species. In the case of Mordecai et al., 2013 fecundity data for *Aedes albopictus* was utilised in place of *Anopheles* fecundity data, with the assumption that the fecundity thermal response of the two genera are not significantly different. If the analysis reveals that there is a significant difference this can be taken into account when evaluating the model. If the responses between genus/species are similar then it means that the data which can be used is expanded, increasing the pool of available data for model parametrisation and evaluation, as well as determining the applicability of model predictions to different vector genus/species.

The temperature dependence of traits may vary as different aspects of vector biology perform best at various temperatures. Knowing the trait distribution can also be useful in identifying the relative importance of a trait at a given temperature, e.g. a trait with higher than average optimum will have a larger contribution to vector dynamics at higher temperatures than other traits.

Materials and Methods

Data Collection

Four separate data sets were gathered from three groups:

- Sadie Jane Ryan, Department of Geography & Emerging Pathogens Institute, University of Florida
- Leah R. Johnson, Integrative Biology, University of South Florida
- Erin Mordecai, Department of Biology, Stanford University

These four data sets were collected from laboratory experiments for various reasons, but were digitised in a common format from the biotraits database (Dell, Pawar & Savage, 2013), as part of an ongoing effort to digitise vector temperature response data. They were either extracted from published studies using special software, or the original data was requested from authors. Due to time constraints no new data was added.

Of the data series gathered non-temperature dependant data was identified and discarded. Each data series was assigned a unique identity, data series were separated based on differences in location, gender, trait, genus, species and paper citation. Standardising the data involved converting trait responses like 'total development time' into rates by taking the inverse (e.g. 1/time). Fractional data was converted into decimal. Trait units were standardised into metric. The format was edited to include the original data, pre-standardisation. Data series that were originally separated but are from the same experiment were joined into a single data series. Traits measuring the same thing were identified and assigned a single common name, e.g. eggs/female/day was converted EFD. Traits descriptions were standardised between for data series of the same trait.

Fitting models

All modelling and data analysis was done using R (version 3.1.1) or Excel 2013.

For each data series was fitted with the polynomial model shown below, at three different levels: 1st order (linear), 2nd order (quadratic) and 3rd order (cubic).

$$y = y_0 + x_1 T + x_2 T^2 + x_3 T^3$$

Plots for each data series with the fitted models were created, an example of which is shown in supp. Figure 1, and the plots for all data series is also available in the supplementary material. The best model was determined using the Akaike information criterion for small sample size (AICc), which was calculated using the AICc function from the R package AICcmodavg (version 2.0-3).

Coefficients from the best fitted models were used to create polynomial models using Polynom (version 1.3-8), from which model summary data was generated, and the values of the stationary points and zeros (x-intercept) were extracted. The optimum temperature was derived by taking the temperature of the stationary point (where dy/dx=0) that corresponds to the maximum trait value. For the data series with an inverse relationship (e.g. mortality) the temperature corresponding to the minimum trait value was taken instead. Maximum and minimum estimated temperatures were the maximum and minimum x-intercept (y-zero value) derived from the model. In the results for the estimated minimum, maximum and optimum temperatures, only the predictions from the significant (f-stat p-value ≤ 0.05) best fitted quadratic or cubic models were used. Estimates from linear best models were ignored.

Optimal measured temperature was identified by finding maximum trait value within a data series and identifying the corresponding temperature. For inverse relationships temperature

corresponding to the minimum trait value was taken instead. The minimum measured temperature is the minimum temperature of a data point within the data series, and similarly for the maximum.

Analysis

Summary of data structure and quality were produced in excel using pivot charts and tables, which were then modified.

Due to the wide variety of traits measured and the lack of common naming, the size of trait data sets were too small for analysis. So results data were put into four broad categories of traits:

- Development: % emergence and development rates for all stages and genders.
- Fertility: gonotrophic cycle length, various measures of the number of eggs laid and fecundity.
- Survival: larval survival, egg to adult survival, daily survival, longevity, lifespan and mortality data which is in essence still a measure of survival (survival probability is 1-mortality probability).
- Transmission: % that become infected and transmit, EIP (day virus detected in mosquito), probability of infection, % infection, % dissemination and blood meals taken per female.

Those that do not belong to any of the four categories (metabolic and water loss rate) were left blank.

There was a lack of measured critical minimum and maximum temperatures, as few of the data series have a wide enough range of temperature such that trait expression reaches zero. Instead the minimum and maximum temperatures measured for each data series is used as a proxy for the thermal limits, with the assumption that they are approximately equal to the true critical thermal maxima and minima.

For data quality histogram plots were produced using ggplot2 (version 1.0.0), a binwidth of 1 was used for the number of data points, number of unique temperatures and range of temperatures, a binwindth of 2 for the minimum and maximum measured temperature.

In the summary table all count data and temperature means were calculated using pivot tables in excel. For a visual representation of thermal responses ggplot2 (version 1.0.0), was used to show the minimum, maximum and optimum measured/estimated temperature, for the four broad trait categories, four genera, two *Glossina* species and two *Aedes* species.

To include all the dependant variables a Multivariate Analysis of Variance (MANOVA) was used to test for the significance of differences between thermal response for traits, genus and species. Format was y ~ x, where x is trait/genus/species and y = Min measured/estimated temperature, Max measured/estimated temperature, Optimal measured/estimated temperature.

Results

Data Structure

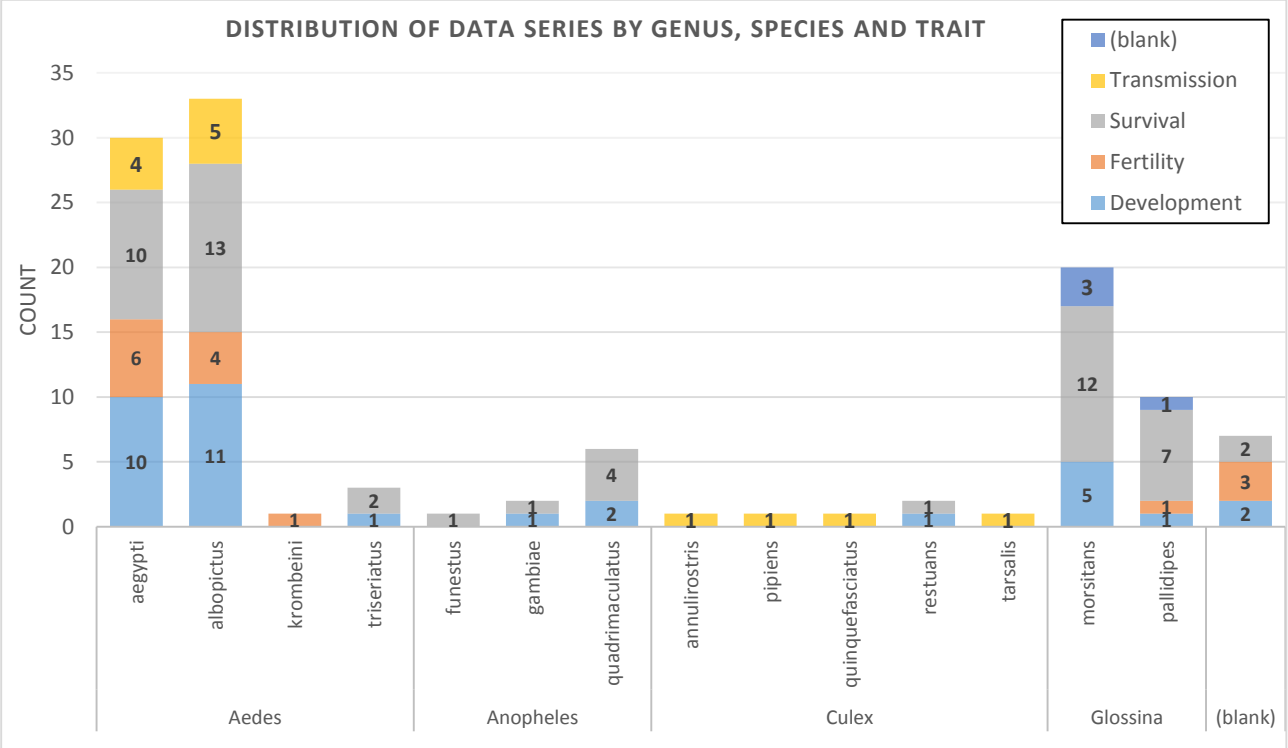

Figure 1 Shows a breakdown of the 119 data series into the numbers present for each genus and species. With blank being those for unknown vectors, or those that do not fit into a broad trait category.

In total there were 119 separate data series. A detailed breakdown of distribution is shown in Table 1, and is visually represented in Figure 2. Within the four broad trait categories survival data accounts for nearly half (45%) of the data, together with development data they make up nearly three-quarters of the data (73%). The vector data is made up of four genera, three of which are mosquitos and one is the Tsetse fly. Within the different genus *Aedes* accounts for more than half of all the data (56%), and the vast majority (82%) of the mosquito data, with *Glossina* making up another quarter (25%) of all the data. At the species level 53% of all the data and 80% of the mosquito data is made up of two species *alone, Aedes aegypti* and *Aedes albopictus*. In addition some data series were not general thermal responses, but looked instead at the thermal response at the critical maximum and minimum temperature.

Trait	No. of Data Series (no. of sig. best model)	No. of Polynomial Order of Best Model (no. of significant best model)			Average of the Measured Temperature			Average of the Estiamted Temperature		
		Linear	Quadratic	Cubic	Minimum	Optimum	Maximum	Minimum	Optimum	Maximum
Development	34 (14)	10 (7)	15 (1)	9 (6)	16.74	29.57	32.43	13.58	31.13	45.09
Fertility	15 (4)	5 (1)	5 (2)	5 (1)	17.24	26.23	32.72	15.50	26.33	40.12
Survival	53 (29)	10 (3)	30 (16)	13 (10)	18.46	25.43	32.01	13.00	23.94	36.61
Transmission	13 (6)	4 (3)	4 (1)	5(2)	16.42	24.50	31.63	7.45	28.07	37.34
(blank)	4 (2)	3 (2)	1 (0)	0 (0)	NA	NA	NA	NA	NA	NA
Total	119 (55)	32 (16)	55 (20)	32 (19)	17.22	26.43	32.20	12.80	25.59	37.24

Genus Species	No. of Data Series (no. of sig. best model)	No. of Polynomial Order of Best Model (no. of sig. best model)			Average of the Measured Temperature			Average of the Estiamted Temperature		
		Linear	Quadratic	Cubic	Minimum	Optimum	Maximum	Minimum	Optimum	Maximum
Aedes	**67 (23)**	**14 (5)**	**39 (14)**	**14 (4)**	**16.29**	**26.16**	**32.94**	**5.56**	**25.10**	**43.18**
aegypti	30 (13)	7 (3)	18 (6)	5 (4)	17.25	27.52	32.67	10.22	25.63	37.44
albopictus	33 (6)	6 (1)	18 (5)	9 (0)	16.42	25.07	32.67	-1.21	25.17	51.55
krombeini	1 (1)	0 (0)	1 (1)	0 (0)	14.00	26.00	35.00	14.67	24.70	34.73
triseriatus	3 (3)	1 (1)	2 (2)	0 (0)	6.00	24.00	38.00	7.00	22.75	38.51
Anopheles	**9 (3)**	**6 (3)**	**3 (0)**	**0 (0)**	**27.01**	**36.04**	**37.16**	**NA**	**NA**	**NA**
funestus	1 (0)	1 (0)	0 (0)	0 (0)	38.00	38.50	40.00	NA	NA	NA
gambiae	2 (2)	2 (2)	0 (0)	0 (0)	31.50	35.76	37.90	NA	NA	NA
quadrimaculatus	6 (1)	3 (1)	3 (0)	0 (0)	23.68	35.67	36.44	NA	NA	NA
Culex	**6 (2)**	**3 (2)**	**2 (0)**	**1 (0)**	**17.00**	**24.53**	**30.92**	**NA**	**NA**	**NA**
annulirostris	1 (0)	1 (0)	0 (0)	0 (0)	20.00	29.17	33.50	NA	NA	NA
pipiens	1 (1)	1 (1)	0 (0)	0 (0)	15.00	22.00	32.00	NA	NA	NA
quinquefasciatus	1 (1)	1 (1)	0 (0)	0 (0)	13.00	30.00	30.00	NA	NA	NA
restuans	2 (0)	0 (0)	2 (0)	0 (0)	20.00	20.00	30.00	NA	NA	NA
tarsalis	1 (0)	0 (0)	0 (0)	1(0)	14.00	26.00	30.00	NA	NA	NA
Glossina	**30 (24)**	**6 (4)**	**9 (5)**	**15 (15)**	**18.40**	**24.86**	**28.71**	**18.91**	**26.08**	**32.34**
morsitans	20 (17)	3 (2)	4 (2)	13 (13)	18.35	25.93	29.42	19.88	27.56	31.62
pallidipes	10 (7)	3 (2)	5 (3)	2 (2)	18.49	22.71	27.31	15.34	21.63	35.01
(blank)	7	3(2)	2 (1)	2 (0)	NA	NA	NA	NA	NA	NA
Total	**119 (55)**	**32 (16)**	**55 (20)**	**32 (19)**	**17.64**	**26.53**	**32.13**	**12.80**	**25.59**	**37.24**

Table 1 A summary table of the data and results. Broken down into the trait data and genus, species data.

Note: Significant best models is defined as: best model f-stat p-value ≤ 0.05. Estimated temperatures are only taken from significant best models.

Data Quality

Figure 2 is a breakdown of the various data quality measures used to analyse the data series. Means of (A) number of data points = 8.2, (B) number of unique temperatures = 6.87, (C) range of temperature measured in data series = 14.49, (D) minimum temperature measured in data series = 17.64, (D) maximum temperature measured in data series = 32.13.

It is expected that all data series will have a unimodal response (either 2nd or 3rd order), as trait temperature responses should follow a Boltzmann-Arrhenius relationship (Dell, Pawar & Savage, 2011). Second or first order relationship indicate a bad quality dataset, e.g. due to insufficient range of temperature measured. 73% have a 2nd or 3rd order best model, but less than half (45%) are significant. Overall only a third of the data have a significant 2nd or 3rd order best model.

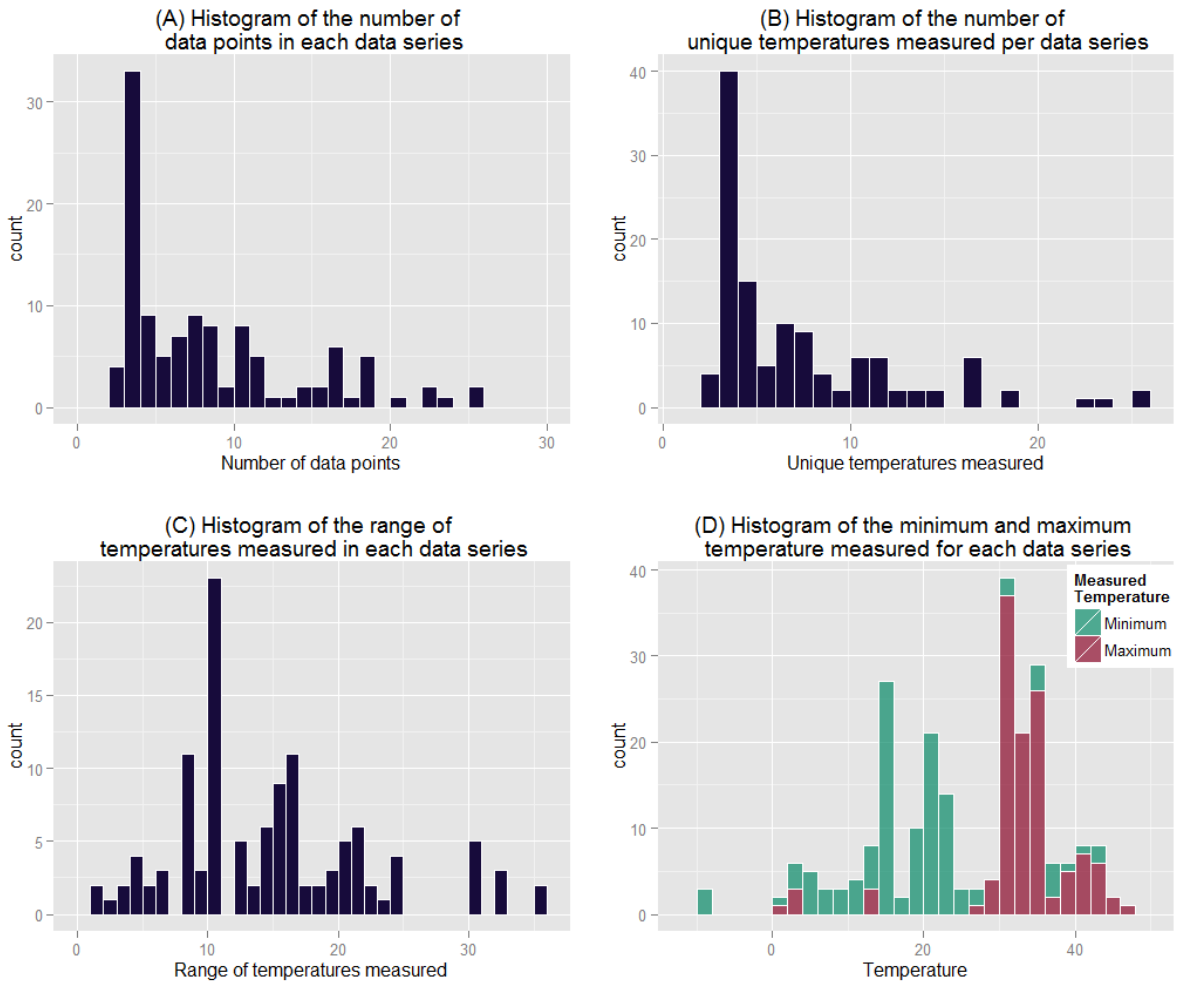

Figure 2 Four histograms displaying various metrics of data quality for the data series present in the unified dataset.

Note: (A) Excludes data series with data points greater than 30, due to outliers created by raw data. Temperature column width (A)(B)(C) = 1 °C, (D) = 2°C.

Analysis

The average temperatures for traits, genus and species is shown in Table 1.

A visual representation of the thermal response of traits is shown in Figure 3. A MANCOVA analysis of the thermal response of trait measured found that there is a significant difference (p-value=0.02117), the analysis using the estimated values (p-value=0.6324) found no significant difference.

Figure 4 shows the temperature response to genus. The MANCOVA of the genus measured temperatures identified a significant difference (p-value=0.000499), with MANCOVA of the estimated values (p-value=0.3971) finding no significant difference.

At the species level Figure 5 shows the thermal responses of *Aedes aegypti* and *Aedes albopictus*, as well as that of *Glossina morsitans* and *Glossina pallidipes.* The other species of the other genus are not shown as they do not have sufficient data to produce meaningful results. The statistical analysis of the variation in thermal response between species of the Aedes genus found no significant difference (p-value for measured=0.1077, p-value for estimated= 0.8434). No significant difference was found in the estimated temperatures for the Glossina genus (p-value=0.5232), but there was a significant difference in the estimated temperatures (p-value=0.02426).

The full results of MANCOVA summary output is shown in Supplementary data Table 1.

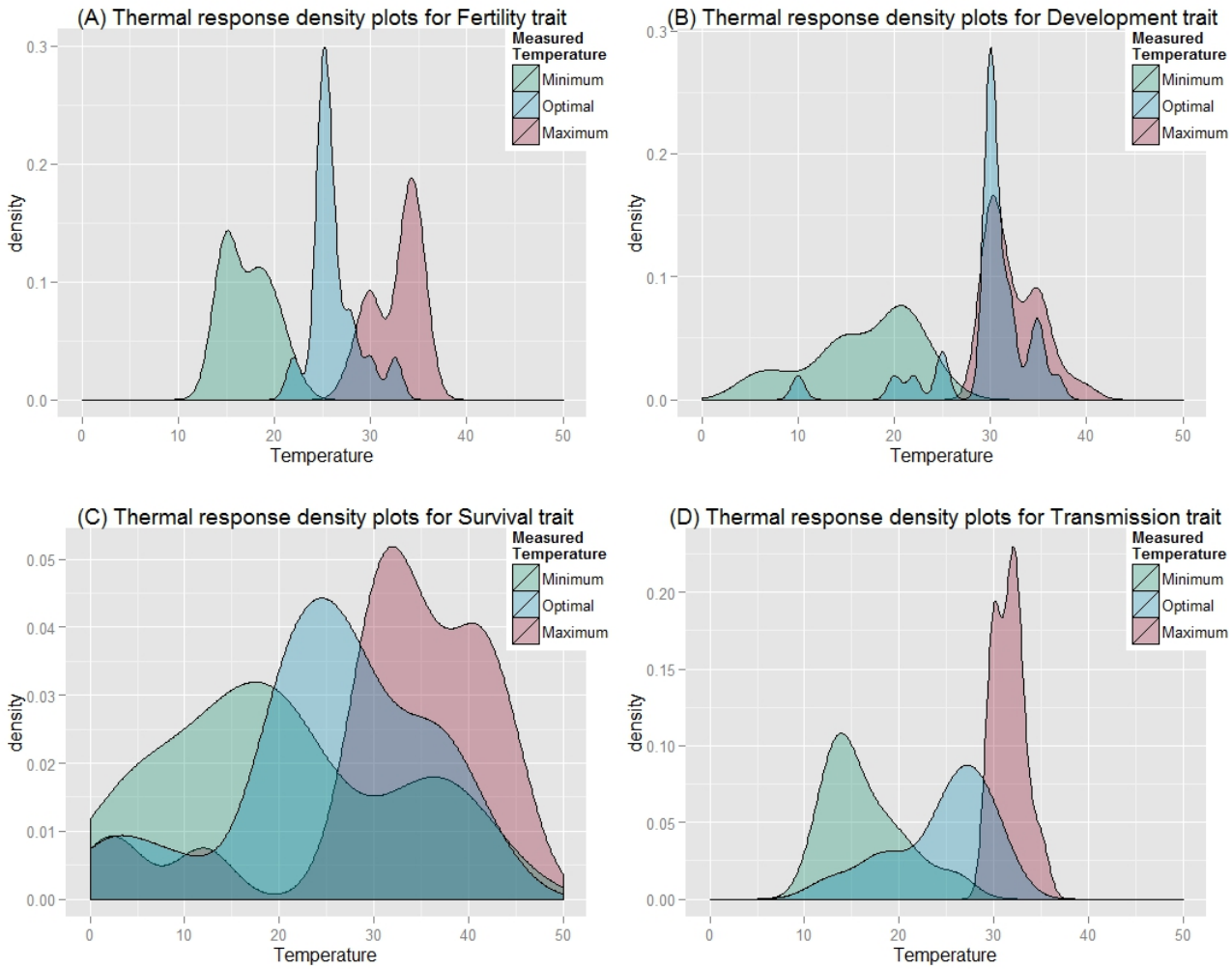

Figure 3 A visual representation of how the thermal response of the four broad traits varies through the location of the peak density and spread of the measured minimum temperature, maximum temperature and optimal temperature.

Note: The minimum and maximum temperatures measured for each data series is used as a proxy for the critical thermal temperatures. Density plot of estimates shown in supplementary Figure 2.

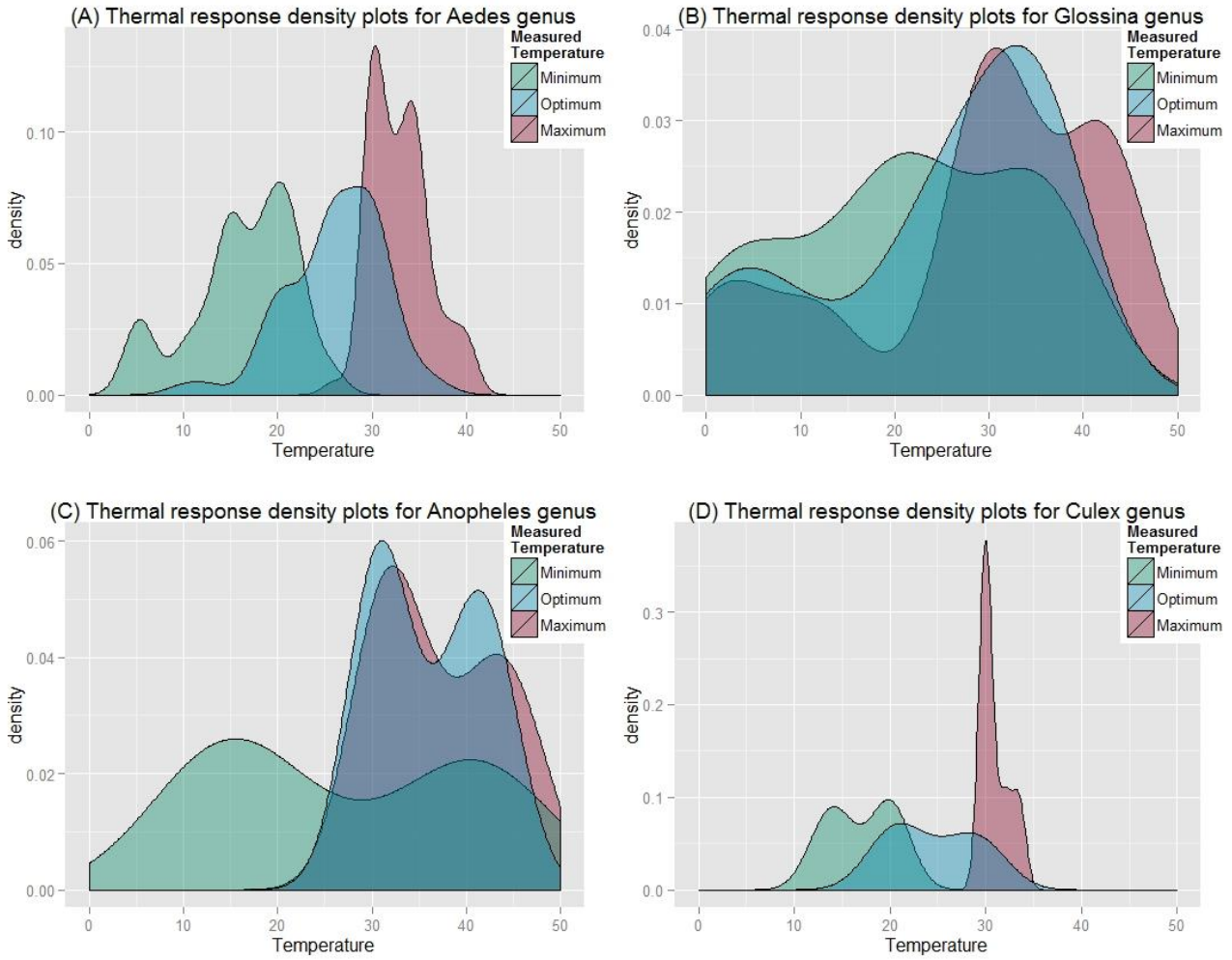

Figure 4 Four density plots are used to show the differences in thermal response between the four genus. Temperature response varies through the location and distribution of the optimum temperature and the critical thermal temperatures.

Note: The minimum and maximum temperatures measured for each data series is used as a proxy for the critical thermal temperatures. Density plot of estimates shown in supplementary Figure 3.

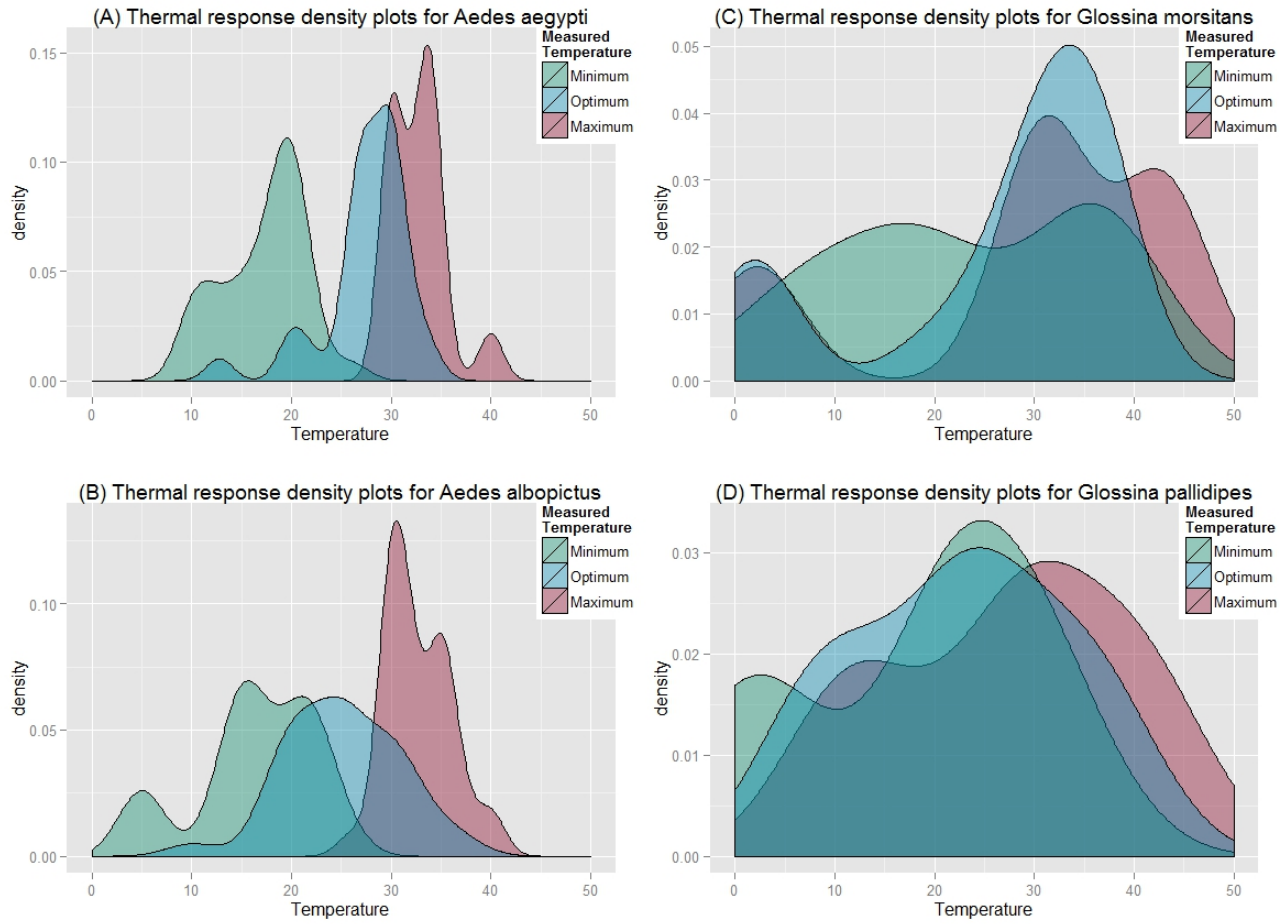

Figure 5 Density plots showing the variation in thermal response at the species level. A&B show two *Aedes* species, whilst C&D show two *Glossina* species. Temperature response varies through the location and distribution of the optimum temperature and the critical thermal temperatures.
Note: The minimum and maximum temperatures measured for each data series is used as a proxy for the critical thermal temperatures. Density plot of estimates shown in supplementary Figure 4.

Discussion

As the results indicate overall the data is very unevenly distributed amongst a few traits, genus and species. This hinders the analysis and means that some of the results from analysis are not likely to be reliable, it also means that the thermal responses may not be accurate representations of the categories as uncertainty is greater with a smaller dataset. It also means that the estimates which rely on good quality of high quality data, are not likely to be accurate predictions. There are other indicators like the cohort size and number of repeats, which was present in some of the collected data, but it was not taken into account when looking at data quality as it is not present for all data series. However in the future this could be incorporated to determine the quality of an individual data series.

As seen in Figure 2 the data quality is overall fairly poor with a low mean number of data points, unique temperatures measured, a small mean range of temperatures and low percentage of data series that significantly fit a unimodal response. Looking at Figure 2(D) less than half the data series have a minimum temperature below the expected critical minimum temperature (≈15°C), similarly less than half measure a maximum temperature past the expected critical maximum temperature

(≈35°C). This means that most of the data series do not capture the full thermal response from end to end. Future studies should aim to have a sufficiently large temperature range, with more unique temperatures than the current mean of 6.87, so that accurate models can be fitted and uncertainty in analysis reduced.

The analysis has many caveats, there is a lack of range of temperature and insufficient unique temperatures for good significant model fits. Density plots of the full thermal responses based on model predictions would be better, however due to lack of significant models sticking with the maximum, minimum and optimum measured temperatures as approximates for the full thermal response will have to do. The broad categories are coarse as it was a way of pooling data to a sufficient size for analysis. Separation into finer categories would be better, as Beck-Johnson et al., 2013 shows that the thermal response varies at different life stages of the *Anopheles* mosquito.

Analysis of the collected dataset has indicated that there is a significant difference between *Aedes* and *Anopheles* genus. However the thermal response of *Aedes* (Table 1, Figure 4A) very closely matches the optimum temperature (≈25°C), critical minimum temperature (≈15°C) and critical maximum temperature (≈35°C) given for *Anopheles* in three other studies (Lunde, Bayoh & Lindtjorn, 2013; Mordecai et al., 2013; Beck-Johnson et al., 2013). This is assuming that the thermal responses of *Aedes aegypti* and *Aedes albopictus* are accurate representations of *Aedes* as a whole, which the species level analysis has indicates to be a valid assumption.

Due to the scarcity of *Anopheles* and *Culex* data within the collected dataset, the analysis between the four genera is unreliable, though we can conclude that there is a significant difference between *Aedes* and *Glossina*, which are completely different insects, and hence the difference would be expected.

The mosquito family is divided into the two subfamilies Anophelinae and Culicinae, the similarity between *Aedes* which belongs to the Culicinae subfamily and *Anopheles* which belongs to Anophelinae-assuming *Aedes* is an accurate representation of the whole subfamily Culicinae-could indicate that the thermal response of the two subfamilies are in fact similar. With more data on the other vector genus of the subfamily Culicinae (*Culex, Culiseta, Haemagogus* and *Ochlerotatus*), analysis can be done to determine if there is significant variation at the genus level of the subfamily. If both subfamilies Anophelinae and Culicinae are similar then all mosquito trait data can be substituted for any disease-vector mosquito genus. In addition model predictions for one genus is likely to be true for other genus as well.

At the species level of comparison, though the statistical analysis is for the whole genus, the lack of data means that we can only really conclude that there is no significant difference in trait response for *Aedes aegypti* and *Aedes albopictus* and that of the two species of *Glossina*. More data is required to test if there is indeed no significant difference between all species of *Aedes*.

These conclusions coupled with genus/species distribution and climate data, could also give an idea about the limits of adaptions to different temperatures. Whereby if the average local environmental temperature is significantly different between *Aedes aegypti* and *Aedes albopictus,* but because the analysis shows there is no significant difference in temperature response, it could mean that there

is a physiological limit to the ability to respond to the long term shifts in temperatures. This sort of information could be useful in understanding temperature responses to a long term shift in temperature.

A more in depth analysis using a mixed effect model should be conducted to test the interaction between traits and genus/species, as trait thermal responses are likely to be adapted to the specific species' environment. However with the current dataset the sample sizes would not be sufficiently large to get accurate results.

Complexity of Vector Dynamics

All the data gathered is lab based, so there remains the question of if it reflects real life accurately. There are many assumptions made in the analysis of the results and in many of the models used. The change in vector dynamics is much more complicated then just temperature alone, with many other factors like socioeconomics, healthcare standards and rainfall all contributing to vector dynamics (Tabachnick, 2010).

A constant temperature laboratory environment does not reflect real life, where there is in fact daily temperature variations. Carrington et al., 2013 shows that temperature variation can have significant effects on traits, especially at high and low temperatures, which is important when looking at the effects of climate change as it acts on the temperature extremes, affecting critical maximum temperature.

Martin & Huey, 2008 predict that the optimal body temperature may in fact be lower than the optimal trait temperature. The right skew of the temperature response curve means that a temperature shift to the right of the optimum produces a larger negative impact on trait expression than an equal shift to the left of the optimum. Hence the equilibrium would be lower than expected. Where the larger the rate of change past the optimum, the lower the true optimum from that of the predicted.

Another assumption made is that there is no adaptability of the vector to changes in temperature. Though not relevant to short term changes e.g. seasonal variation, this is important for climate change predictions that will occur over many years. Studies will need to look into the physiological capacity/limits and rate of adaptation for the various disease vector, and determine if it will have a significant effect on model predictions or not.

Rainfall and humidity has a significant effect on population dynamics of many arthropods, in this case it is controlled for within the laboratory, and is not a limiting factor (Mordecai et al., 2013). However at the same time the future changes in precipitation patterns are much harder to predict and as such model predictions incorporating precipitation may not be accurate (Lunde, Bayoh & Lindtjorn, 2013).

Data acquisition

Due to the small size of the dataset it is unlikely that it is an accurate representation of vector thermal response data as whole, which means that this dataset cannot really be used to inform data acquisition priority.

However in the future should more data be added, the information on data structure can be paired with a sensitivity analysis of the popular models which can determine the relative importance of each trait. The importance can then be matched with a combined quality and number metric to determine which traits warrant greater focus. Though there may be issues with regards to the models used and the variation in sensitivity in addition to the usage of different variables in the models.

Future usage of dataset

There are still many issues with the data that is collected, based on possible errors in the original data collection. Some of the data has a wide range of values suggesting that there may have been treatment differences. Some of the original data had been converted into the factors used for the model, like mosquito development rate that does not clarify which stages the development rates represents. These issues should be sorted out if the dataset is to be used in the future for model parametrisation and evaluation.

This is only data collected from three groups that have closely collaborated, however there are other groups who may have collected data which can be added to the current dataset. Though there is the issue of these groups collecting data only relevant to their use, and may not have included variables like the experimental conditions in the extraction.

When collecting the data together it was brought to my attention that this is the first time that the vector trait data is being brought together from different sources. Bringing together the data had been more difficult than expected and is a labour intensive process, not only in organising the data but collecting it from the different groups. Many studies and models rely on collecting new data, extracting the raw data from the original papers or requesting original data from authors, or parameter values are taken from previous papers where the origin of the value may be lost, and the quality of the original data is uncertain.

Unlike genomics there is a lack of databases that support the growing amount of ecological data gathered. By putting the data collected into a database built on a platform like SQL, could provide the foundation to a web accessible database. The data can be managed by administrators ensuring a common format and quality to the data, which helps with the addition of new data. The database can be kept up to date with contributions from the various groups involved, in addition to making the data retrievable for those who wish to use the data, which also creates the possibility for larger collaborations between many groups.

However there are numerous issues, such as server hosting and administration, which would need to be looked into. There would also need to be an awareness of the database to get contributions to it. The ultimate aim would be that authors would add their own data to the database reducing the need for third parties to extract the information from publications, in addition to providing extra information that was collected but not published. Studying the format and methodology of the large and prevalent genomic databases may provide useful information to how to setup a useful database.

As evidenced by the divergence in the interpretation of the format and data added, more strict guidelines would need to be setup for data collection, so that addition of the data is closer to the intended format reducing the need to reformat the data. Categorisation for traits need to be made more general as well, so that it is not mosquito specific but can include vectors relevant to plants and animals.

References

Beck-Johnson, L. M., Nelson, W. A., Paaijmans, K. P., Read, A. F., Thomas, M. B. & Bjornstad, O. N. (2013) The Effect of Temperature on Anopheles Mosquito Population Dynamics and the Potential for Malaria Transmission. *Plos One.* 8 (11), e79276.

Brown, J., Gillooly, J., Allen, A., Savage, V. & West, G. (2004) Toward a metabolic theory of ecology. *Ecology.* 85 (7), 1771-1789.

Campbell, L. P., Luther, C., Moo-Llanes, D., Ramsey, J. M., Danis-Lozano, R. & Peterson, A. T. (2015) Climate change influences on global distributions of dengue and chikungunya virus vectors. *Philosophical Transactions of the Royal Society B-Biological Sciences.* 370 (1665), 20140135.

Carrington, L. B., Armijos, M. V., Lambrechts, L., Barker, C. M. & Scott, T. W. (2013) Effects of Fluctuating Daily Temperatures at Critical Thermal Extremes on Aedes aegypti Life-History Traits. *Plos One.* 8 (3), e58824.

Dell, A. I., Pawar, S. & Savage, V. M. (2013) The thermal dependence of biological traits. *Ecology.* 94 (5), 1205-1206.

Dell, A. I., Pawar, S. & Savage, V. M. (2011) Systematic variation in the temperature dependence of physiological and ecological traits. *Proceedings of the National Academy of Sciences.* 108 (26), 10591-10596.

Gillooly, J., Brown, J., West, G., Savage, V. & Charnov, E. (2001) Effects of size and temperature on metabolic rate. *Science.* 293 (5538), 2248-2251.

Githeko, A., Lindsay, S., Confalonieri, U. & Patz, J. (2000) Climate change and vector-borne diseases: a regional analysis. *Bulletin of the World Health Organization.* 78 (9), 1136-1147.

Johnson, L. R., Ben-Horin, T., Lafferty, K. D., McNally, A., Mordecai, E., Paaijmans, K. P., Pawar, S. & Ryan, S. J. (2015) Understanding uncertainty in temperature effects on vector-borne disease: a Bayesian approach. *Ecology.* 96 (1), 203-213.

Lunde, T. M., Bayoh, M. N. & Lindtjorn, B. (2013) How malaria models relate temperature to malaria transmission. *Parasites & Vectors.* 6, 1-10.

Martin, T. L. & Huey, R. B. (2008) Why "suboptimal" is optimal: Jensen's inequality and ectotherm thermal preferences. *The American Naturalist.* 171 (3), E102-18.

Mordecai, E. A., Paaijmans, K. P., Johnson, L. R., Balzer, C., Ben-Horin, T., de Moor, E., McNally, A., Pawar, S., Ryan, S. J., Smith, T. C. & Lafferty, K. D. (2013) Optimal temperature for malaria transmission is dramatically lower than previously predicted. *Ecology Letters.* 16 (1), 22-30.

Tabachnick, W. J. (2010) Challenges in predicting climate and environmental effects on vector-borne disease episystems in a changing world. *Journal of Experimental Biology.* 213 (6), 946-954.

Supplementary Material

The URL (https://www.dropbox.com/sh/rrznayez5o63me8/AABmKixnL17bB7yyY9TtSP92a?dl=0) contains the following material:

- Full combined dataset
- Original datasets
- Results from analysis
- Code used for analysis
- Individual data series plots with fitted models

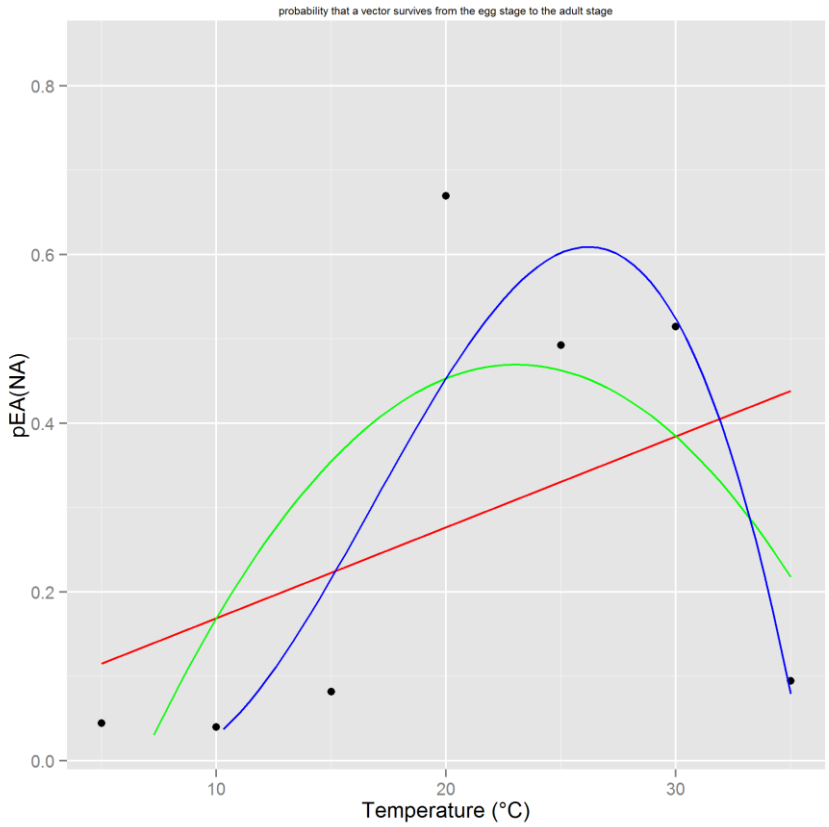

Figure 1 Plot illustrating three models fitted to data where: Red = linear, Green = quadratic, Blue = cubic. Shows how the quadratic model takes into account the unimodal shape, with the cubic allowing greater flexibility in fit, in this case showing the expected sharp decline past the optimum temperature.

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Trait Measured
             Df  Pillai approx F num Df den Df  Pr(>F)
Broad.Trait   4 0.20636   2.0314     12    330 0.02117 *
Residuals   110

Genus Measured
              Df  Pillai approx F num Df den Df   Pr(>F)
Vector.Genus   4 0.29627   3.0134     12    330 0.000499 ***
Residuals    110


Trait Estimated
            Df  Pillai approx F num Df den Df Pr(>F)
Broad.Trait  3 0.27809  0.78327      9     69 0.6324
Residuals   23

Genus Estimated
             Df  Pillai approx F num Df den Df Pr(>F)
Vector.Genus  2 0.24396   1.0651      6     46 0.3971
Residuals    24

Aedes Measured
               Df  Pillai approx F num Df den Df Pr(>F)
Vector.Species  3 0.22697    1.637      9    180 0.1077
Residuals      60

Glossina Measured
               Df  Pillai approx F num Df den Df Pr(>F)
Vector.Species  1 0.08125  0.76644      3     26 0.5232
Residuals      28

Glossina Estimated
               Df  Pillai approx F num Df den Df  Pr(>F)
Vector.Species  1 0.59404   4.8776      3     10 0.02426 *
Residuals      12

Aedes Estimated
               Df  Pillai approx F num Df den Df Pr(>F)
Vector.Species  3 0.49186  0.52295      9     24 0.8434
Residuals       8
```

Table 1 Collected summary results from MANOVA analysis. Where y = Min measured/estimated temperature, Max measured/estimated temperature, Optimal measured/estimated temperature.

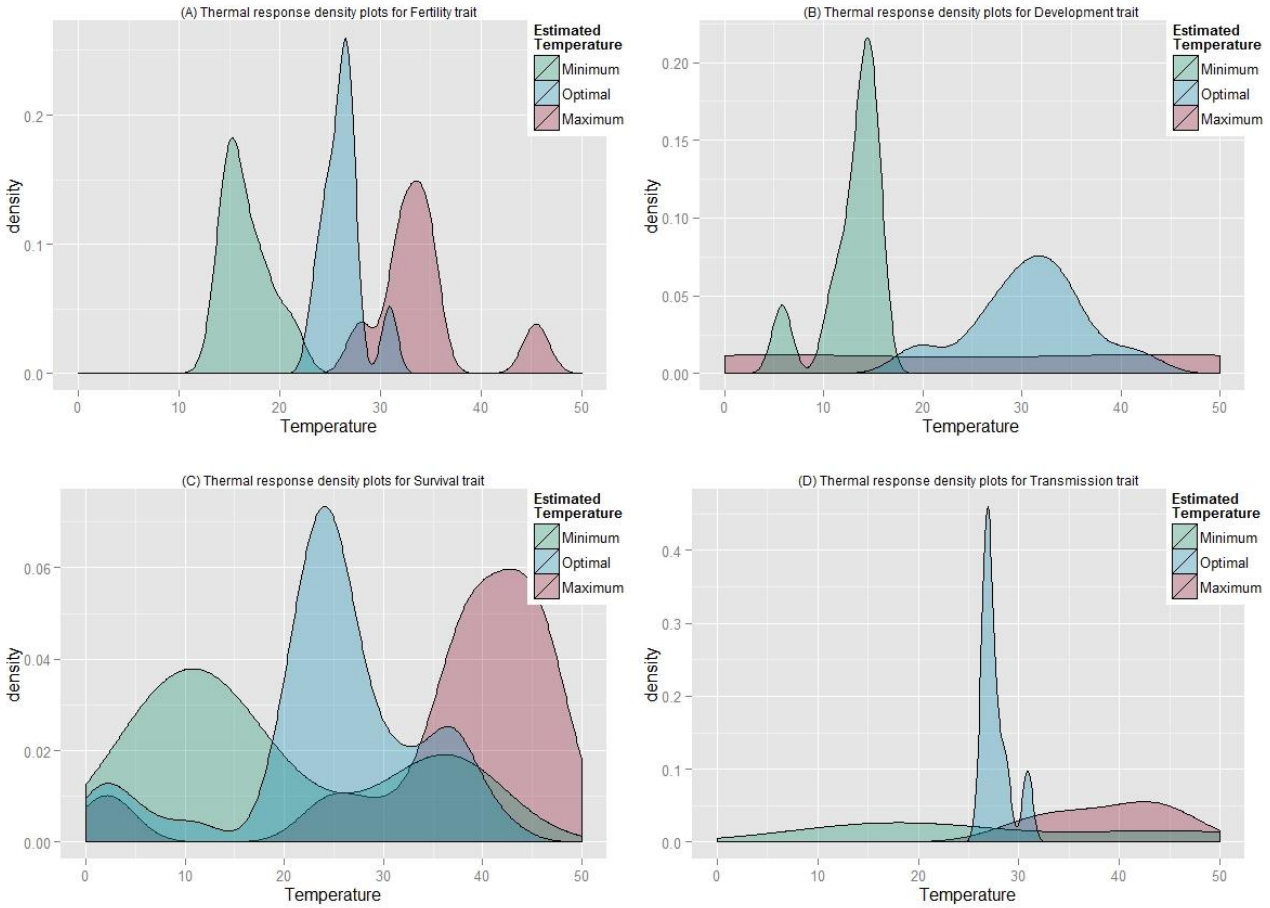

Figure 2 Graphical representation of the predicted thermal responses of the four broad trait categories, based on estimates from significant unimodal models.

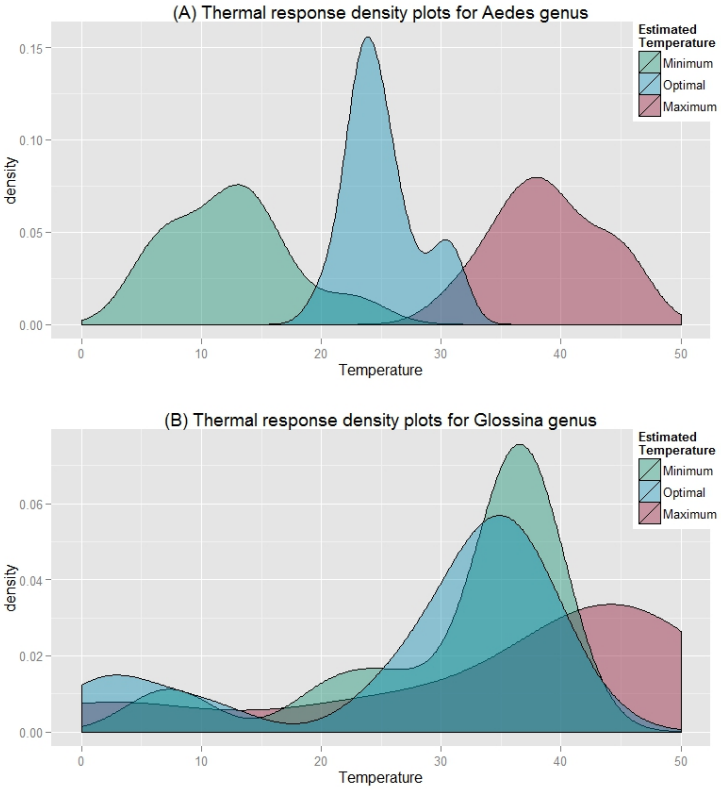

Figure 3 Graphical representation of the predicted thermal responses between the two genus with sufficient data for estimates, based on estimates from significant unimodal models.

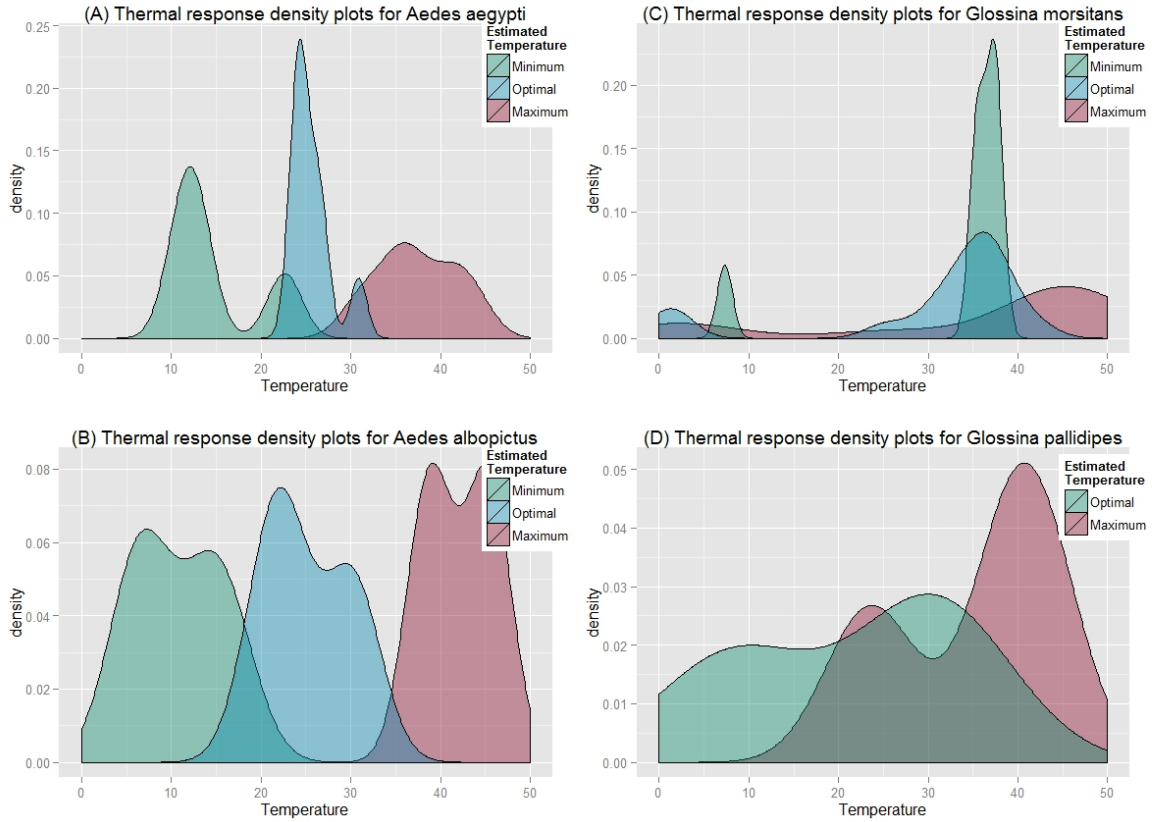

Figure 4 Graphical representation of the predicted thermal responses between two *Aedes* species (A&B) and two *Glossina* species (C&D), based on estimates from significant unimodal models