**High temperature structures tropical forest *Drosophila* communities: evidence from physiological and population level measures**

**Results**

The nine *Drosophila* species shown in figure 1a accounted for 99% of samples. Abundance patterns across the altitude gradient differed among species. The regression coefficient evaluated how detection probability change along with altitude (supplementary figure 4). Values of the regression coefficients correlated tightly with the weighted central altitude (hIndex) of distribution (rho = 0.98, p-value < 0.001, Spearman’s rank test). For simplification, the hIndex was used to represent the distribution pattern in the following analysis.

*D. bipectinate* and *D. pandora* were categorized as lowland-biased species with high confidence. *D. pseudoananassae* was most likely to bias towards lowland. *D. rubida* and *D. sulfurigaster* showed no significant change with altitude. *D. birchii* was most likely to bias towards highland. *D. palidifrons* and *D. pseudotakahashii* were significantly enriched in high altitudes. The only six samples of *D. bunnanda* were all found at low altitude, which might explain the peculiar value of its estimated coefficient (coefficient = -69, not shown in figure 1b) and its large standard error (se = 21603). Another larger-scale study also confirm that it was found predominantly at lowland (Schiffer and McEvey 2006). Therefore, *D. bunnanda* was categorized as lowland-biased species. As shown by supplemental figure XX, species that are closely related did not share the same distribution type.

Thermal performance curves of daily fecundity per female vary in the range, optimal temperature, height, and shape among species (figure 2). Table 1 shows estimates of the parameters of the Briere’s function for each species. *D. bunnanda* and *D. birchii* are closely related, but have completely opposite distribution types. The lowland-biased species *D. bunnanda* have higher minimal temperature, optimal temperature and maximal temperature than the upland-biased *D. birchii*, corresponding to their distribution types. In contrast, *D. sulfurigaster* always outperforms its upland-biased relative, *D. palidifrons.* The temperature for optimal reproductive performance (RTopt) did not correlate with their distribution patterns (coefficient = 0.068, 95% ci = -1.93 – 2.07)

Values of RTmin had no relationship with the species distribution patterns (coefficient = 0.024, 95% ci = -2.47 – 2.52). Similarly, upland-biased species did not show higher fecundity at the stressfully-low temperature, 17°C (p = 0.788). For example, *D. palidifron* has the highest RTmin and lowest reproductive success at 17°C, while it is found predominantly at high altitude. When exposed to acute sublethal low temperature (5°C), except for *D. simulans* and *D. melanogaster*, the other *Drosophila* species all show similarly weak resistance (male: p = 0.18; female: p = 0.53). All species recovered their fecundity after eight-day exposure to 14°C. This recovered fecundity showed a minor but not significant increase (coefficient = 0.35, p = 0.105) for the upland-biased species. They also spent a longer time to recover from the chill coma (male: p = 0.054; female: p = 0.029), which is presumably disadvantaged in the upland environment. Species that had lower RTmins did not show any advantage in the short-term cold resistance or cold recovery (spearman’s rank test, see table 2).

Species whose distribution were biased towards lowland show higher RTmax (coefficient = -2.52, 95% ci = -3.68 - 1.36, p = 0.00125). Reproductive performance at 29°C decrease with hIndex (coefficient = -5.09, p < 0.0001). After exposure to 29°C for eight days, neither of the two highland-biased species could reproduce when transferred back to mild temperature. Five out of the six non-biased and lowland-biased species resumed reproduction. Heat tolerance, measured by knockdown time at high temperature, was lower among species whose distribution were biased towards high latitude (male: coefficient = -9.1, p = 0.0013; female: coefficient = -5.4, p = 0.056). Species which had higher RTmax also stayed active for longer in the extreme high temperature (spearman’s rank test, see table 2).

Values of RTmin have more variation among species than the RTmax, as predicted by the “climate variability hypothesis”. Heat tolerance was not found to have a trade-off relationship with cold tolerance: RTmin and RTmax were not significantly correlated (supplementary figure 6); (on the contrary) Species that resisted heat for longer also showed stronger tolerance to cold, indicated by longer resistance to chill coma and shorter recovery time.

**Figure 1. a) Proportion of samples found in the low-, middle- and high-altitude site for the nine Drosophila species. b) Regression coefficients and hIndex unanimously describe altitudinal distribution patterns. *D. bunnanda* is not included in the graph because its regression coefficient and standard error are peculiarly large in absolute value due to its small sample size. Error bars show 90% confidence intervals.**

A close up of a logo

Description automatically generated

A screenshot of a cell phone

Description automatically generated

**Figure 2. The thermal performance curve of reproduction. Color is ordered by their distribution pattern, with highland-biased species labeled by cold color and lowland-biased species labeled by warm color.**

A close up of a device

Description automatically generated

**Figure 3. Reproduction in stressful temperature treatment and after the temperature treatment. The species on the horizontal axis are ordered descendingly by the mean relative altitude of their distribution (hIndex). hIndex of *D. simulans* and *D. melanogaster* is unavailable. Samples of *D. simulans* were only found in lowland sites, so it was placed after *D. bunnanda*. *D. melanogaster* was not found in our study site, but used as an internal control.**

A close up of text on a white surface

Description automatically generated

**Figure 4. Physiological responses to lethal heat stress and cold stress. The species on the horizontal axis are ordered descendingly by the mean relative altitude of their distribution (hIndex). hIndex of *D. simulans* and *D. melanogaster* is unavailable. Measurements of females were labeled by red, males were labeled by blue.**

**A screenshot of a cell phone

Description automatically generated**

**Table 1 Estimated parameters of thermal performance functions and their 90% credible intervals (CI) of the nine species.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **species** | **a** | **CI\_a** | **b** | **CI\_b** | **RTmin** | **CI\_RTmin** | **RTmax** | **CI\_RTmax** |
| *D. bipectinata* | 0.0046 | 0.0030 - 0.0059 | 1.26 | 1.01 - 1.55 | 15.28 | 14.56 - 15.88 | 30.45 | 30.08 - 31.05 |
| *D. birchii* | 0.0034 | 0.0022 - 0.0056 | 1.17 | 0.95 - 1.57 | 13.45 | 13.08 - 13.79 | 29.25 | 28.11 - 29.80 |
| *D. bunnanda* | 0.0017 | 0.0012 - 0.0026 | 0.88 | 0.81 - 1.07 | 14.58 | 14.09 - 15.20 | 31.19 | 30.61 - 31.77 |
| *D. melanogaster* | 0.0037 | 0.0032 - 0.0042 | 1.72 | 1.48 - 2.02 | 8.32 | 6.93 - 9.38 | 32.13 | 32.03 - 32.28 |
| *D. palidifrons* | 0.0073 | 0.0055 - 0.0099 | 1.74 | 1.36 - 2.39 | 16.23 | 15.51 - 16.77 | 29.07 | 28.14 - 29.39 |
| *D. pandora* | 0.0052 | 0.0037 - 0.0065 | 1.25 | 1.03 - 1.51 | 15.26 | 14.56 - 15.79 | 30.13 | 29.88 - 30.57 |
| *D. pseudoananassae* | 0.0053 | 0.0035 - 0.0071 | 1.67 | 1.22 - 2.33 | 15.07 | 14.15 - 15.91 | 29.22 | 28.42 - 29.80 |
| *D. simulans* | 0.0035 | 0.0027 - 0.0047 | 1.68 | 1.36 - 2.22 | 8.51 | 6.94 - 9.66 | 31.08 | 30.38 - 31.78 |
| *D. sulfurigaster* | 0.0040 | 0.0027 - 0.0051 | 1.26 | 1.03 - 1.53 | 14.37 | 13.92 - 14.94 | 30.12 | 29.84 - 30.63 |

**Table 2. Correlation matrix among thermal traits (RTmin, RTmax, RTopt, female knockdown time to heat, male knockdown time to heat, female knockdown time to cold, male knockdown time to cold, female recovery time from cold, male recovery time from cold). Spearman’s rank correlation rho is shown in the table. Significant correlation (p < 0.05) is labeled as bold.**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | RTmin | RTmax | RTopt | FKDHEAT | MKDHEAT | FKDCOLD | MKDCOLD | FRCCOLD | MRCCOLD |
| RTmin | 1.00 | -0.55 | 0.02 | **-0.89** | -0.53 | -0.31 | -0.56 | 0.53 | 0.65 |
| RTmax |  | 1.00 | 0.17 | **0.71** | **0.93** | 0.64 | 0.61 | **-0.88** | **-0.85** |
| RTopt |  |  | 1.00 | -0.01 | 0.33 | 0.63 | 0.54 | -0.41 | -0.30 |
| FKDHEAT |  |  |  | 1.00 | **0.72** | 0.53 | **0.71** | -0.57 | **-0.68** |
| MKDHEAT |  |  |  |  | 1.00 | **0.79** | **0.71** | **-0.89** | **-0.83** |
| FKDCOLD |  |  |  |  |  | 1.00 | **0.85** | **-0.74** | **-0.75** |
| MKDCOLD |  |  |  |  |  |  | 1.00 | **-0.72** | **-0.70** |
| FRCCOLD |  |  |  |  |  |  |  | 1.00 | **0.93** |
| MRCCOLD |  |  |  |  |  |  |  |  | 1.00 |

Supplementary figure 1. The change of fecundity during 1st – 2nd day to 7th – 8th day in different temperature.

A close up of a whiteboard

Description automatically generated

Supplementary figure 2. Time table of fecundity measurement.

A screenshot of a cell phone

Description automatically generated

Supplementary figures 3. Diagnostics of model fitting of thermal performance curve.

A screenshot of a cell phone

Description automatically generatedA close up of a white wall

Description automatically generatedA close up of a map

Description automatically generated

Supplementary figure 4. Examples of logistic regression on occurrence data. *D. bipectinata* is lowland-biased species. *D. rubida* shows no bias. *D. palidifrons* is highland-biased species.

A screenshot of a video game

Description automatically generated

Supplementary figure 5. Daily fecundity and fitted thermal performance curve of each of the nine species.

A bunch of different colors

Description automatically generated

Supplementary figure 6. Scatter plot of posterior samples of RTmin and RTmax parameters.

A screenshot of a cell phone

Description automatically generated

Supplementary figure 6. Daily temperature of Feb. against the lowest CTmax and highest CTmin.

Chart, application

Description automatically generated

**Supplementary table 1. Isofemale line used to construct MBLs.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Species | Origin of cultured lines (yes/no) | | | | Lines for MBLs |
| Kirrama low1 | Kirrama high1 | Paluma low1 | Paluma high1 |
| *D. bunnanda* | yes | no | yes | no | KL87, KL134, KL127, PL114 |
| *D. pandora* | no | no | yes | no | PL17, PL21, PL012 |
| *D. bipectinata* | yes | no | yes | no | KL84, KL43, PL85, PL20 |
| *D. pseudoananassae* | yes | yes | yes | no | KL19, KH25, PL30, KH42 |
| *D. sulfurigaster* | yes | yes | yes | yes | KL08, KH10, PL51, PH18 |
| *D. rubida* | yes | yes | yes | yes | Construction unfinished3. |
| *D. birchii* | yes | yes | yes | yes | KL22, KH26, PL122, PH169 |
| *D. palidifrons* | no | yes | no | yes | KH20, KH69, PH183, PH184 |
| *D. simulans* | no | no | yes | no | PL45, PL34, PL42, PL43 |
| *D. pseudotakahashii* | no | yes | no | yes | Did not construct MBLs4 |

Note:

1. “Low” means sites from low altitude. “High” means sites from high altitude.
2. The three isofemale lines were the only lines cultured at the lab.
3. Construction is not finished by the start of the experiment in May. *D. rubida* grew poorly on the purchased fly medium before changing food recipe and made at the lab. Therefore, its crossing starts later than other species (April 5th). Additionally, *D. rubida* has significantly longer generation time than other species.
4. Only two isofemale lines were cultured at the lab.

**Methods**

1. Study system

Field data were collected, and laboratory cultures were initiated, from rainforest sites spanning elevations of from 59 – 916 m at Paluma Range (S18° 59.031' E146° 14.096') and Kirrama Range (S18° 12.134' E145° 53.102'), Queensland, Australia. Mean temperatures at study sites ranged from 21°C to 26°C.

*Drosophila* isofemale cultures were established in 2017 and 2018 from pupae collected from both high- and low-altitude sites. Cultures had been maintained at 24°C and 12/12 L/D cycle at the Biology Centre, Czech Academy of Sciences since collection and transferred and maintained at 25°C and 12/12 L/D cycle at the Department of Zoology, the University of Oxford, UK, since December 2018. Theywere maintained for approximately 15 to 30 non-overlapping generations in Czech Republic and additional approximately four to seven non-overlapping generations in Oxford before they were used to construct mass bred lines (see below).

1. Field distribution survey

*Drosophila* pupae were sampled using bottle traps baited with fermented banana from 11th March – 12th April 2016. Details were described in Jeffs et al. 2020. Relative abundance was surveyed for three sites representing the highest, lowest, and most central points of each of the two mountains. 182 pupae were sampled at each site. 716 pupae were successfully identified to species by DNA metabarcoding (Jeffs et al. 2020), with 86 – 134 pupae at each site. *D. serrata* (1 individual) and *D. immigrans* (4 individuals) were excluded from the distribution analysis due to infrequent occurrence.

1. Preparation of experimental animals

To revive genetic variation, we constructed mass bred lines (MBLs) by combining four isofemale lines of each *Drosophila* species (except for *D. pandora*, where only three isofemale lines were available). The four lines were selected from different mountains and different altitude if possible (Detailed arrangement was shown in supplementary table 1). Each population cage was initiated using two independently-reared MBLs of the same species. Cages were maintained at 25°C and 12/12 L/D cycle for more than four generations before the experiment. Therefore, thermal traits should not have been influenced by maternal effect, acclimation, or isofemale line effect. MBLs of eight Australian *Drosophila* species and one laboratory strain (wild typeDah) of *D. melanogaster* were used for laboratory measurement of thermal performance. *D. melanogaster* does not occur naturally at the study sites, was measured together with the focal species as a benchmark for future comparisons.

Fly eggs collected from the population cage were reared under low-density (less than 100 eggs per vial) at 25°C and 12/12 L/D cycle. Within 12 hours of emergence, virgin females and males were separately kept at 25°C and 12/12 L/D cycle. We additionally mixed five females and five males in each of two vials to monitor their reproductive activity every day. As different species have different development times and sexual maturation times, eggs of different species were collected from cages on different days so that their first day of egg-laying was synchronized. Two days after sexual maturation, half of the adults were subjected to fecundity measurement and the other half, which were siblings, were subjected to physiological measurement.

1. Fecundity measurement

Two virgin females were paired with two virgin males on a 4ml *Drosophila* medium (Percentage concentration (weight/volume): 8% corn flour, 4% yeast, 5% sugar, 1% agar, and 1.67% methyl-4-hydroxybenzoate.). They were randomly subjected to water baths set at one of the seven constant temperature (14°C, 17°C, 20°C, 23°C, 26°C, 29°C, 32°C) and 12/12 L/D cycle. Vials were submerged in water baths. The water level was kept above the area that flies could freely move. The temperature and humidity of each water bath were monitored in two additional empty tubes. The level of humidity was similar to field condition, ranging between 80% - 95%. The observed temperature showed ±0.5°C fluctuation around the mean temperature, which was used as the corrected temperature in analysis (Supplementary table 2 [both centre and corner logs]).

As fecundity changed through time and this trend of change was influenced by temperature (Supplementary figure 1), the offspring numbers were measured for the 1st – 2nd day and the 7th – 8th day and they were combined to reflect relative fecundity. After eight-day temperature treatments, all flies were kept at 25°C for another four days to examine the recovery of reproduction. Surviving flies were recorded at the beginning and end of each period. Vials containing eggs produced during the testing periods were maintained at their corresponding temperature for development. Vials were examined daily for emergence. The first emergence dates were recorded for different temperatures and the numbers of F1 adults were frozen after 5-7 days and counted later. For each species and each temperature treatment, eight replicates were evenly split between two blocks. The detailed schedule is shown in supplementary figure 2.

1. Physiological measurement

Critical temperatures (CTmin and CTmax), half lethal temperature (LT50cold and LT50heat), heat stress survival, knockdown (heat coma and chill coma) time, recovery time, etc. have been common practice to compare relative resistance to heat stress and cold stress (Gibert et al. 2001; Hoffmann et al. 2003). Critical temperature becomes more popular for its direct association of climate data. However, the absolute value of CTmin and CTmax is significantly influenced by the rate of temperature change (Terblanche et al. 2007), making it difficult to compare between studies. Knockdown time and recovery time are simpler measure which also serve the purpose of comparing thermal tolerance. Variation has been observed among species and geographic ranges (Gibert et al. 2001; Hoffmann, Anderson, and Hallas 2002), and knockdown/recovery time were correlated with other measurement, such as critical temperature (Andersen et al. 2015).

How to assess Drosophila heat tolerance: Unifying static and dynamic tolerance assays to predict heat distribution limits

knockdown time has linear relationship with testing temperature (static). They accord well with distribution ecotypes. Knockdown time at 40C capture the variance and produce time scale which is easy to measure logistically. (Jørgensen et al., 2019)

Relative tolerance to extreme cold temperature was measured by individuals’ knockdown time at 5°C and the time for recovery of mobility after a 30-minute exposure to 5°C. Constant temperature for cold stress are often chosen around 0°C (Gibert et al. 2001). As tropical species often have significantly lower cold resistance (Gibert et al. 2001), 5°C was used instead to increase the variation among the tested species (pilot results not shown). Heat stress is chosen to be 40°C, following common practice (Hoffmann et al. 2003). After being knocked down by heat (40°C), most flies did not survive. In this case, only knockdown time was used to evaluate tolerance to the extremely high temperature.

[Measures of CTmin, LTe50 and LTi50 proved to be the best predictors to describe the variation in realized latitudinal distributions. there was only a weak correlation between the entrance into coma (CTmin) and the recovery from chill coma (CCRT) (Anderson, 2015)]

Virgin adult flies were kept in groups at 25°C and 12/12 L/D cycle for 9-10 days before randomly-selected individuals were allocated separately in empty flat-bottom 3ml insect tubes. An observation rack was divided into 3X3 grids and each grid held seven tubes containing the same species. Nine species were assigned in random order to one grid. The observation rack was moved immediately into the incubator, representing the start of the heat or cold treatment. Every tube was examined once every minute and the flies that lost or recovered their motor ability were recorded. For each sex, three identical blocks of the above procedures were repeated.

1. Data analysis

All statistics were performed with R statistical software (version 3.6.0 ). All analysis code is available in ??GITHUB or SUPP??. *D. melanogaster* and *D. simulans* were not included in analyses involving field distribution, because their distribution patterns were unavailable.

*Distribution*. To calculate the abundance-weighted central altitude (hIndex) of distribution, the relative location of each sample was assigned 0, 0.5, and 1 if it was collected at low-, middle- and high-altitude sites. In addition, intra-specific abundance patterns were assessed by logistic regression of detection probability with altitude. For each species, pupal identity was labeled as 1 if the pupa was identified as the focal species and 0 if it was identified as any other species. Their identity was fitted against the altitude as the fixed effect and the transect as the random effect in generalized linear mix-effect model (varying intercept, varying slope) using the *lmer* package. These two ways of describing distribution patterns were compared using a Spearman’s rank test.

It is important to note that the abundance patterns we focus here is to compare the abundance of a species with itself along the altitude, rather than to compare the abundance of multiple species in a given location. By this definition, an upland-biased species may have higher absolute value of population size than the lowland-biased species in the lowland.

*Thermal performance curve*. A multi-level, non-linear piecewise model was fitted under the Bayesian framework using MCMC sampling within the *rstan* package in R. Total offspring numbers were calculated by combining the offspring numbers on day 1-2 and day 7-8. The average daily fecundity per female was calculated, then square root transformed. Square-rooted daily fecundity was modeled with the Briere2 function (Briere et al. 1999):

Y = a \* T \* (T - RTmin) \* (RTmax - T)^(1/b) (RTmin < T < RTmax),

Y = 0 (T <= RTmin or R >= RTmax),

where T is the temperature, *RTmin* and *RTmax* is the minimum and maximum temperature for the species to reproduce, *a* is a scaling factor and *b* is a shape factor of the curve. *a*, *b*, *RTmin*, and *RTmax* of the nine species were assumed to share normal distribution respectively. Square rooted daily fecundity was modeled using a normal distribution with temperature-dependent standard deviation. Assuming temperature dependency of standard deviation generated better fitting than assuming the same standard deviation across temperature treatments, as judged by leave-one-out cross-validation. It was because when the temperature was equal or close to *RTmin* and *RTmax*, the standard deviation should be zero or close to zero. A normal distribution is not ideal to model the transformed count data, which are all positive. However, modeling offspring counts with Poisson, zero-inflated Poisson, negative binomial, lognormal distribution did not produce converged results, potentially due to the piecewise nature of the thermal performance function. Diagnostics were performed and the model performance is acceptable (supplementary figure 3). Non-informative priors were chosen for all parameters. The values of *a* were bounded to be positive. The values of *b* were bounded to be larger than 0.8 to ensure that the thermal performance curve has a steeper slope on the right side. The values of *RTmin* were bounded to be lower than 17°C and the values of *RTmax* were bounded between 26°C - 35°C according to experience.

Median of the posterior distribution of *a*, *b*, *RTmin*, and *RTmax* were used as the model parameters to construct the thermal performance curve.

The model parameters were also estimated by the maximum likelihood method using the *bbmle* package. Total offspring numbers were modeled by Poisson distribution. This method was not multi-level; therefore, the shapes of curves of different species varied more than when assuming shared distributions of model parameters. Besides, this method behaved badly in estimating the uncertainty of the parameter estimation. Nevertheless, the ranks of the *RTmax* (rho = 0.88, p = 0.003, Spearman’s rank test) and *RTmin* (rho = 0.97, p = 0.00016, Spearman’s rank test) estimated by both methods are highly correlated.

*Reproduction-related traits*: The posterior distributions of *RTmin*, *RTmax*, and *RTopt* (6000 samples of each parameter of each species) were modeled by hIndex as the fixed effect and species identity as the random effect in the linear mix-effect model. Fecundity of 29°C and 17°C, recovered fecundity after 29°C and 14°C were used as direct measurements of their performance in the high and low temperatures. The offspring numbers were modeled by hIndex and experimental block as fix effects and species as a random effect in the generalized linear mix-effect model (family = “negative binomial”). Diagnostics of the models were conducted. Data points with extreme leverage value were excluded and the model was fitted again to test if the statistical significance still holds.

*Physiological tolerance*. The six measurements of physiological tolerance were modeled by hIndex, block, and tube position as fixed effects, species as a random effect in the linear mix-effect model.

*Correlation analysis*. The pairwise correlation among thermal traits was evaluated by Spearman’s rank test in *Hmisc* package. Traits included RTmin, RTmax, RTopt, median knockdown time to hot of female (FKDHEAT), median knockdown time to hot of male (MKDHEAT), median knockdown time to cold of female (FKDCOLD), median knockdown time to cold of male (MKDCOLD), median recovery time from cold of female (FRCCOLD), median recovery time from cold of male (MKDCOLD).