Temperature control of larval dispersal and the implications for marine ecology, evolution, and conservation

Mary I. O'Connor*, John F. Bruno, Steven D. Gaines, Benjamin S. Halpern, Sarah E. Lester, Brian P. Kinlan, Jack M. Weiss

*To whom correspondence may be addressed: email: maryo@unc.edu

SI TEXT

Supporting Methods.

Additional detail on data compilation. We searched BIOSIS and ISI Web of Science online databases for relevant articles and identified additional articles using their cited references. Our goal was not to find all relevant articles, but to build a database of sufficient size and taxonomic diversity to test our hypothesis. When multiple studies for a single species met our criteria, we chose the study that reported a greater number of unique test temperatures.

We used the following criteria to include data: (*i*) studies reported planktonic larval duration (*PLD*) (hatching to metamorphosis or settlement) at multiple temperatures for a single species, (*ii*) other environmental factors (e.g., salinity, food availability) were constant and within the species' normal range, (*iii*) temperatures were nonlethal, and (*iv*) the interval between maximum and minimum test temperatures exceeded 1.5 °C. We extracted the following information for each species: temperature treatments, mean planktonic larval duration at each temperature (we used the minimum or median value when necessary), normal temperature range for the species, geographic location of collection, and larval size.

Incorporation of larval mass. Data on larval mass were sparse. Larval size at metamorphosis was most commonly available as maximum or standard larval length (mm), and for this data set values range over three orders of magnitude (0.2–26.8 mm). There is no convincing or standard method for converting larval length to volume for most species, so a quantitative estimate of the potential effect of mass on *PLD* requires further research.

Potential effects of larval size on model selection (Results). Larval size could modify the temperature dependence of *PLD* in two ways. Systematic bias would occur if larvae tested at different temperatures were collected from different regions and had adapted their size to native temperature conditions (1). The studies we included tested the effect of temperature on larvae collected at a single location, so this kind of counter-gradient variation does not occur in our dataset. Variation in size with temperature could also occur due to phenotypic plasticity resulting in larger larvae at colder temperature treatments (2). Of 17 studies in our analysis that tested for an effect of temperature on larval size, only seven report a significant effect. Therefore, for a subset of species in this analysis, temperature may affect *PLD* indirectly through its effects on size in addition to the direct effects of temperature.

Temperature measurement error as a source of error in our model (Discussion). Because these studies are peer-reviewed, we consider measurement error to be minimal and evenly distributed among species. There are two types of measurement error that could influence variance in *PLD* about the regression line for any particular species. One type is thermometer error, which we assume to be relatively minor. A possibly more important source of variation is fluctuation in treatment temperature between measurement times. This could occur if, for example, temperature fluctuated overnight but measurements were always made during the day.

Composite multilevel model including mean ln(temperature) and developmental mode as predictors. This is the multilevel model referred to in Fig. 4 and SI Tables 6 and 7.

Level 1: $\ln(PLD)_{ij} = \beta_{0i} + \beta_1 * (\ln(T_{ij} / \mathrm{T_c})) + \beta_2 * (\ln(T_{ij} / \mathrm{T_c}))^2 + \varepsilon_{ij}$, $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$

Level 2: $\beta_{0i} = \beta_3 + \beta_3 * (\text{Mean } \ln(T_i)) + \beta_4 * (\textit{developmental mode}_i) + u_{0i}$, **[7]**

$u_{0i} \sim \mathrm{N}(0, \tau^2)$

Percent variance explained by mean ln(temperature) (Pseudo R-squared). In multilevel models there is no direct equivalent to the R^2 of ordinary linear regression. We calculated a pseudo-R^2 value as described by Singer and Willett (3). This value estimates the amount of variance explained by a model relative to a null model of the same form. We entered mean ln(*temperature*) in our model as a level-2 predictor for the intercept, β_{0i}. The pseudo-R^2 we calculate here measures the reduction in the intercept variance, τ^2, that occurs when mean ln(temperature) and developmental mode are added to the model given in Eq. **4** to produce Eq. **7**.

Predicting the Trajectories of New Species. One of the attractive features of multilevel models is that they typically outperform classical regression in predictive accuracy (4, 5). Using the multilevel model we've developed one can directly predict the *PLD* of any species used in building the model at a temperature that was not observed. With additional data (one or more observations) it is also possible to use this model to predict the *PLD* of a marine species that was not among those used in estimating the model. The validity of such a prediction hinges on whether the temperature model we've proposed is truly universal. In this section we explain how to use our model to make predictions for new species.

The exponential-quadratic model we've proposed is shown in its generic composite form below.

$$\ln(PLD)_{ij} = \beta_0 + u_{0i} + \beta_1 * (\ln T_{ij} - \ln \mathrm{T_c}) + \beta_2 * (\ln T_{ij} - \ln \mathrm{T_c})^2 + \varepsilon_{ij} ,$$

where $u_{0i} \sim N\left(0, \tau^2\right)$ and $\varepsilon_{ij} \sim N\left(0, \sigma^2\right)$. Based on this the average PLD–temperature trajectory for species *i* is the following:

$$\ln(PLD)_{ij} = \beta_0 + u_{0i} + \beta_1 * (\ln T_{ij} - \ln \mathrm{T_c}) + \beta_2 * (\ln T_{ij} - \ln \mathrm{T_c})^2$$

or, written as a level-1 model

$$\ln(PLD)_{ij} = \beta_{0i} + \beta_1 * (\ln T_{ij} - \ln \mathrm{T_c}) + \beta_2 * (\ln T_{ij} - \ln \mathrm{T_c})^2, \qquad \textbf{[8]}$$

where $\beta_{0i} = \beta_0 + u_{0i}$. The terms β_0, β_1, and β_2 are fixed effects while u_{0i} is a random effect that is unique for species *i*. Before Eq. **8** can be used to make predictions for a new species, the random effect for that species needs to be estimated from data. Because random effects are random variables rather than population parameters, it is more correct to speak of "predicting" random effects rather than "estimating" them, a convention that we adopt here. Our discussion of the prediction of random effects closely parallels Fitzmaurice *et al.* (2004), pp. 206–210 (6).

From Eq. **8** we see that predicting u_{0i} is equivalent to predicting β_{0i} so we'll focus on this latter quantity instead. It can be shown that the best linear unbiased predictor (BLUP) of β_{0i} is a linear combination of the population-averaged estimate $\hat{\beta}_0$ and the ordinary least squares estimate $\hat{\beta}_0^{OLS}$.

$$\hat{\beta}_{0i} = w_i * \hat{\beta}_0^{OLS} + \left(1 - w_i\right) * \hat{\beta}_0. \qquad \textbf{[9]}$$

The OLS estimate $\hat{\beta}_0^{OLS}$ is the ordinary regression estimate obtained by using only the n_i observations available for species *i* while $\hat{\beta}_0$ is the population estimate obtained from the multilevel model we present using all the available species. The weight w_i appearing in Eq. **9** is a ratio of the between species variability to the sum of the within- and between-species variabilities. These quantities are listed in SI Table 2. For a mixed model with a single random effect (i.e., Eq. **8**) this ratio is the following.

$$w_i = \frac{n_i * \tau^2}{n_i * \tau^2 + \sigma^2}, \qquad \textbf{[10]}$$

where again n_i is the number of temperature observations available for species *i.* To use eqn S4 we substitute estimates of τ^2 and σ^2 obtained from the multilevel model. Generally REML (restricted maximum likelihood) estimates of variance components are preferred over maximum likelihood (ML) estimates for this purpose because they tend to be less biased. For our data the differences between the REML and ML estimates of the variance components turn out to be fairly small, so we don't bother further with this distinction in what follows.

$\hat{\beta}_{0i}$ as given by Eq. **9** is called a shrinkage estimate because it causes the OLS estimate $\hat{\beta}_0^{OLS}$ to be more or less "shrunk" toward the population-averaged value $\hat{\beta}_0$ depending on the relative magnitudes of τ^2, σ^2, and the sample size n_i of the full, multi-species database. If most of the variability in the data occurs between species then w_i assigns more weight to the OLS estimate. If on the other hand within species variability is dominant, the shrinkage estimate will more closely resemble the population-averaged value. Observe that there is no restriction on size of n_i; it can be as small as one. Just as with τ^2, larger values of n_i will move the shrinkage estimate closer to the OLS estimate.

We illustrate the methodology with an example. Suppose we have the following data for the chiton *Tonicella lineata* that consists of two observations at two different temperatures.

Table 1. Data for new species

Temperature	PLD
10.0	3.83
12.5	2.75

This species actually occurs in our database and was used in fitting the model, but for point of illustration we'll treat it as a new species. This will also allow us to check our work. Parameter estimates from the exponential-quadratic multilevel model obtained using all 69 species and $T_c = 15$ are the following.

Table S2. Parameter estimates

Parameter	Estimate	Standard error
β_0	3.1671863	0.10692115
β_1	– 1.3439341	0.04640600
β_2	– 0.2775613	0.04096706
τ^2	0.75299956	—
σ^2	0.02296691	—

To make use of Eq. **9** we need $\hat{\beta}_0^{OLS}$. To obtain it we first solve for β_{0i} in Eq. **8**.

$$\beta_{0i} = \ln(PLD)_{ij} - \beta_1 * \left(\ln T_{ij} - \ln T_c\right) - \beta_2 * \left(\ln T_{ij} - \ln T_c\right)^2 \qquad \textbf{[11]}$$

Using SI Table 1 and the parameter estimates from SI Table 2, Eq. **11** yields the following two separate estimates of β_{0i}.

```
mod1<-lme(log(PLD) ~ I(log(temp)-log(15))+I((log(temp)-log(15))^2),
     random=~1|species,data=inverts.red,method=‘ML’)
b0i.func<-function(x) log(x[2])-fixef(mod1)[2]*(log(x[1])-log(15))-
     fixef(mod1)[3]*(log(x[1])-log(15))^2
b0i.func(c(10,3.83))
I(log(temp) - log(15))
              0.843578
b0i.func(c(12.5,2.75))
I(log(temp) - log(15))
             0.7757992
```

The ordinary least squares estimate of β_{0i} is just the mean of these two values. Note: If SI Table 1 consisted of only one temperature observation for the new species then the OLS estimate would be the single value obtained from applying Eq. **11** to this one observation.

```
b0.OLS<-mean(c(b0i.func(c(10,3.83)),b0i.func(c(12.5,2.75))))
b0.OLS
[1] 0.8096886
```

The term $\hat{\beta}_0$ needed in Eq. **9** is just the population-averaged value reported in Table S2. Next we calculate the weights.

```
as.numeric(VarCorr(mod1)[1,1])->tau2
as.numeric(VarCorr(mod1)[2,1])->s2
wi<-2*tau2/(2*tau2+s2)
```

Finally we can apply Eq. **9**.

```
wi*b0.OLS+(1-wi)*fixef(mod1)[1]
(Intercept)
   0.845101
```

This is the same answer returned by R that is obtained when we extract the fixed and random effects and sum the results.

```
ranef(mod1)[[1]][67]+fixef(mod1)[1]
(Intercept)
   0.845101
```

Because between-species variability dominates the within-species variability for our data, i.e. $\tau^2 >> \sigma^2$, the shrinkage estimate ends up being very close to the OLS estimate even though we have only two temperature observations for the new species.

The individual trajectory (empirical Bayes estimate) for *Tonicella lineata* is the following.

$$\ln(PLD)_{ij} = 0.845101 - 1.3439341 * \left(\ln T_{ij} - \ln 15\right) - 0.2775613 * \left(\ln T_{ij} - \ln 15\right)^2$$

We can also write the trajectory in the uncentered form,

$$\ln(PLD)_{ij} = \gamma_0 + \gamma_1 * \ln T_{ij} + \gamma_2 * \left(\ln T_{ij}\right)^2,$$

by multiplying things out and grouping terms. Alternatively, the following transformation equations can be used.

$$\begin{aligned}\gamma_0 &= \beta_0 - \beta_1 * \ln 15 + \beta_2 * (\ln 15)^2 \\ \gamma_1 &= \beta_1 - 2\beta_2 * \ln 15 \\ \gamma_2 &= \beta_2\end{aligned}$$

For our example the transformation equations yield the following uncentered equation for *Tonicella lineata*, predicted from the model based on the multi-species database reported here.

$$\ln(PLD)_{ij} = 2.449036 + 0.1593659 * \ln T_{ij} - 0.2775613 * \left(\ln T_{ij}\right)^2$$

References

1. Conover, DO & Schultz, ET (1995) Trends in Ecology & Evolution 10, 248-252.
2. Atkinson, D (1994) Advances in ecological research 25, 1-58.
3. Singer, JD & Willett, JB (2003) Applied longitudinal data analysis (Oxford University Press, New York).
4. Ashartous, D & de Leeuw, J (2005) Journal of Educational and Behavioral Statistics 30, 109-139.
5. Gelman, A (2006) Technometrics 48, 432-435.
6. Fitzmaurice, GM, Laird, NM, & Ware, JH (2004) Applied Longitudinal Data Analysis (Wiley, New York).
7. Brown, JH, Gillooly, JF, Allen, AP, Savage, VM, & West, GB (2004) Ecology 85, 1771-1789.
8. Gillooly, JF, Brown, JH, West, GB, Savage, VM, & Charnov, EL (2001) Science 293, 2248-2251.
9. Burnham, KP & Anderson, DR (2002) Model selection and multimodel inference (Springer-Verlag, New York).
10. Verbeke, G & Molenbergs, G (2000) Linear mixed models for longitudinal data (Springer-Verlag, New York).