Ecography

ECOG-04725
Giménez, L., Exton, M., Spitzner, F., Meth, R., Ecker, U., Jungblut, S., Harzsch, S., Saborowski, R. and Torres, G. 2020. Exploring larval phenology as predictor for range expansion in an invasive species. – Ecography doi: 10.1111/ecog.04725

Supplementary material

Appendix 1. Source sink dynamics

We simulated a life cycle where juveniles and adults distribute in discrete habitat patches (Fig. A1a). The model resembles that presented by Armsworth (2002) but ours lacks age structure and contains a spatial component enabling variation in larval survival along the spatial gradient. Adults reproduce in local habitats with fecundity (*f*) and release gametes or larvae to the water column. Larval dispersal is driven by the coefficients (d_{ij}) of the dispersal matrix (D): larvae may either return to the natal habitat, colonise a new habitat or die through overdispersion. During dispersal, larvae survive in the pelagic habitat with probability = L_{ij} due to source of mortality other than overdispersion. Larvae settling in a habitat patch, metamorphose and survive the juvenile phase with probability = J_p.

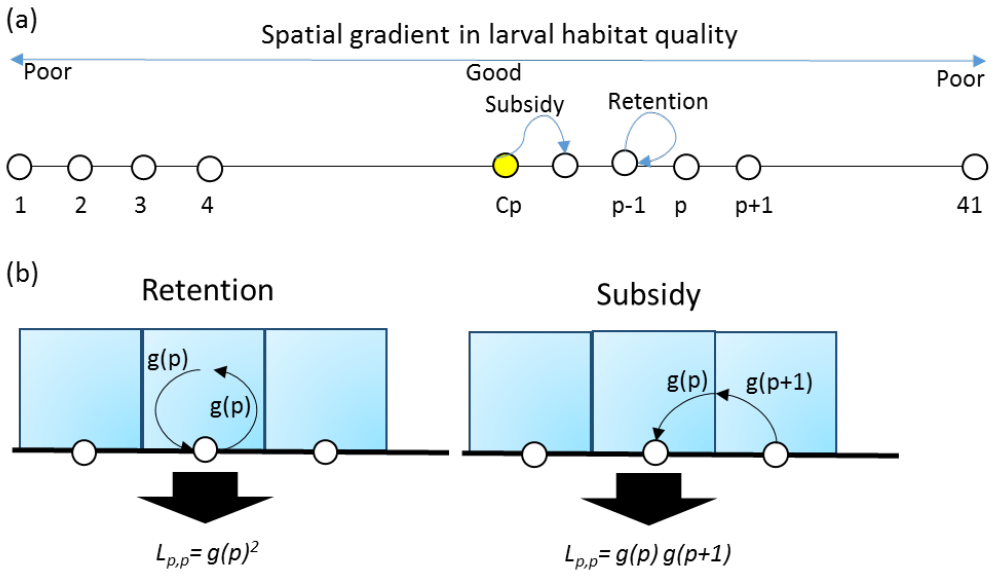

Figure A1. Components of the metapopulation model. (a) The metapopulation is distributed in 41 habitat patches (circles) along a linear shore. Along the shoreline there is a gradient in habitat quality from the centre (good quality) toward the sides (poor quality). At the centre, there is an occupied patch (in yellow); at time = t_0 all remaining patches are empty. Any patch, p, is connected only to itself (retention) and the adjacent patches (subsidy). (b) Larval survival: the pelagic habitat is modelled as collection of coastal cells (in blue) associated to each habitat patch (circles). Any given cell, p, is characterised by its quality, which drives larval survival. Larval dispersal is modelled as a two-step processes (indicated by arrows) of equal duration and dispersal distance, each with survival probability g(p). Retention: larvae leave and return to the original patch (as indicated by arrows) remaining in the same coastal cell; such larvae survive with probability $g(p)^2$. Subsidy: larvae leave a habitat patch (e.g., p+1), cross two coastal cells and settle in an adjacent patch, p; such larvae will survive with probability $g(p)\cdot g(p+1)$.

The dynamics of the metapopulation is modelled in each of 41 habitat patches as a discrete process where the abundance of a given population, p, is computed by the equation:

$$N_{(t+1,p)} = N_{(t,p)} \cdot f \cdot d_{p,p} \cdot L_{p,p} \cdot J_p + N_{(t,p-1)} \cdot f \cdot d_{p,p-1} \cdot L_{p,p-1} \cdot J_p + N_{(t,p+1)} \cdot f \cdot d_{p,p-1} \cdot L_{p,p+1} \cdot J_p \text{ (eq. 1)}$$

The right hand side of eq. 1 has three terms driven by retention (first term) and subsidy from the populations located immediately adjacent to the target population (i.e. other terms of the connectivity matrix are zero).

Juvenile survival is density-dependent following the Beverton-Holt model:

$$J_p = 1/(1+\beta S_p), \text{ (eq. 2)}$$

$$\text{with } S_p = f \cdot [N_{(t,p)} \cdot d_{p,p} \cdot L_{p,p} + N_{(t,p-1)} \cdot d_{p,p-1} \cdot L_{p,p-1} + N_{(t,p+1)} \cdot d_{p,p-1} \cdot L_{p,p+1}] \text{ (eq. 3)}$$

Each patch has an associated pelagic coastal cell (Fig. A1b) characterised by a specific quality. Variation in the quality of the coastal cell drives larval survival according to a function, *g(i)*. Larval survival modelled as two-step process, each characterised by a value of *g(i)* depending on the position of the larvae in the coastal cell. For retained larvae, both steps occur in the same cell; hence, for population *P*, the survival probability is computed as $L_{p,p} = g(p)^2$ (i.e. independence is assumed). For larvae arriving from adjacent populations, each step occurs in a different coastal cell. Therefore, survival is computed as:

$$L_{p,p-1} = g(p) \cdot g(p-1) \text{ (eq. 4a)}$$

$$L_{p,p+1} = g(p) \cdot g(p+1) \text{ (eq. 4b)}.$$

The generating function is given as $g(i) = \{\exp[-(i-P_C)^2/2\sigma^2]\}^{0.5}$ where *Pc* is the population located at the centre of distribution and σ is a constant controlling the rate of spatial decay in survival. This function ensures that the survival probability (as driven by the quality of the pelagic coastal cell) is maximal at P_C and decreases toward the borders. In all simulations $f = 10$, $\beta = 0.01$ and $\sigma = 6$ are constant.

In the first step, the model is run for 40 time-steps to produce Figure 1b, with $N_0 = 10$ at P_C and zero in any other population. In the second step, the connectivity matrix has only retention coefficients (= 0.34) and the model is run for additional 40 times, using the abundance patterns obtained at the 40th iteration in step 1. In Figure 1c, local extinction at each population p was defined when $N_p < 0.1$.

The model was run as a matrix model using R and the library *matrixcalc*. We created a spatial projection matrix out of three square matrices (dispersal, fecundity and survival matrices) with rows and columns corresponding to each of the local populations. The dispersal matrix, D, contains the retention coefficients (= 0.34 for all diagonal elements) and additional coefficients (dij = 0.33 for all sub and supra-diagonal elements and dij =0 otherwise) denoting the dispersal

to and from adjacent populations. (2) The fecundity matrix, F, is a diagonal matrix with values (f=10 for all diagonal elements). (3) The larval survival matrix, L, contains larval survival parameters placed as in the dispersal matrix (lii = $g(i)^2$, lij = g(i)g(j) in sub and supra-diagonal cells and lij =0 otherwise) . The spatial projection matrix was then computed as:

$$SPM = (D \times F) * L \text{ (eq. 5)}$$

Where $\times$ is the inner product and $*$ is the Hadamard product. The product D $\times$ F gives a matrix where each dispersal coefficient is multiplied by the fecundity so that we have *nij = dij :f*. The Hadamard product gives the terms of SPM as *sii = dii·f·Lii* (diagonal), *sij = dij·f ·Lij* (subdiagonal and supradiagonal) but other terms are *sij = 0* as larvae can only disperse to adjacent cells. The number of larvae arriving to each local habitat (eq. 3) is computed as a vector:

$$S(t) = SPM \times N(t), \text{ (eq. 6)}$$

where N is a vector containing the abundance at each of the 41 local populations. Then, S is used to compute a vector of juvenile survival probabilities J (eq. 2) for each local habitat. Finally, we obtain the vector N(t+1) = S · J, i.e. through element-by-element multiplication.

Reference:

Armsworth, P.R. (2002). Recruitment limitation, population regulation and larval connectivity in reef fish metapopulations. *Ecology 83,*1092-1104.

Appendix 2. Methods: Laboratory experiments

In all experiments, larvae were cultured under controlled conditions using standard methods (see e.g. Torres et al. 2011). Larvae were reared at the different temperatures in separate temperature-controlled rooms (precision ± 0.5°C) set at a constant photoperiod (12:12 hs. light/dark). The rearing vessels were always open to ensure oxygen exchange, and contained 0.2 µm filtered seawater at salinity 32 (checked with a salinometer WTW).The seawater was maintained at each temperature in large containers placed in the temperature-controlled rooms with permanent aeriation. Freshly hatched *Artemia* sp. nauplii were provided *ad libitum* as food at densities ~5 nauplii ml^{-1}. Water and food were changed daily by gently pipetting the larvae out of the rearing vessels, dead specimens were discarded, and live animals were counted in a glass bowl under a binocular microscope, when needed, to distinguish larval stages. Vessels were washed with hot tap water, rinsed with seawater and refilled with appropriate volumes of fully aerated seawater; live individuals were then returned to the vessels.

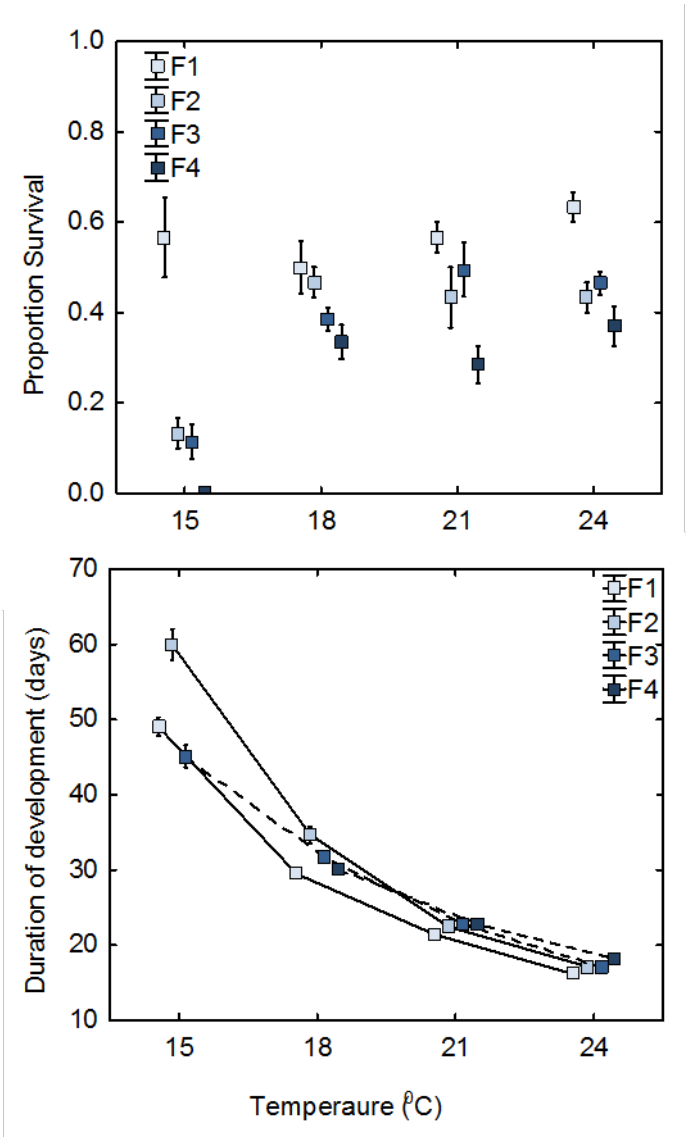

Figure A2. *Hemigrapsus sanguineus*. Maternal variation in survival and duration of development. Larvae from females F1 and F2 were reared in 3 replicate groups of 10 individuals in 100-ml glasses. Larvae from female F3 and F4 were reared in 4 replicate groups of 50 individuals in 500-ml glasses. Accordingly, all larvae were reared at the same starting density of 1 ind*10 ml^{-1}. Error bars are standard error.

Appendix 3. Methods: Temperature time model (TTM)

Our version of the model is coded in R (R core team 2013) using the package lme4 (Bates et al. 2015). The model computes the fraction of development according to a starting day (day of hatching) and the temperature on that given day. On subsequent steps, the fractions are accumulated until they reach 1 (=100% of the larval development from hatching to metamorphosis). At that stage, the duration of development and the timing of predicted metamorphosis to megalopa are recorded and the procedure starts again with a new starting day. The fraction of development was computed by the function *simulate* of lme4, set to simulate new values for all of the random effects. The code used in the simulation is:

```
simulate(modelname, seed=azar, newdata=mynewdata, re.form=NA, allow.new.levels=T)
```

We therefore use the default behaviour (using re.form=NA), which conditions the output on none of the random effects, and simulates new values for all of the random effects (see: https://www.rdocumentation.org/packages/lme4/versions/1.1-20/topics/simulate.merMod).

The model simulates 1,000 events of larval release per hatching day in order to incorporate the random effects in the predictions. However, the model validation is based on the overlap between the predicted and observed timing of megalopa and not on the abundance patterns.

References:

Bates, D. et al. 2015. Fitting linear mixed-effects models using lme4. – J. Stat. Software 67:1–48.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Appendix 4. Methods: Zooplankton techniques

Abundance of zoea I larvae was estimated from plankton samples (oblique tow with HYDROBIOS ring trawl equipped with a Calcofi net, flowmeter and depressor, 500 µm mesh size, 100 cm opening, 4 m length). Samples were taken between June and November 2018, once or twice a week at the Helgoland Roads time series station (54.188° N, 7.900° W) between 7 and 10 am.

142 Zooplankton samples were preserved with sodium tetraborate buffered formaldehyde (4%) and
143 stored in the dark until later taxonomic analyses. For the analyses, each sample was sieved over
144 a 500 µm mesh and rinsed with tap water for 5-10 minutes depending on the zooplankton
145 density. The entire sample was then examined in a Bogorov chamber with a stereomicroscope
146 (Leica M205). Crab zoea l stages belonging to the family Varunidae were removed with forceps
147 and preserved in 70% ethanol for later identification. The remaining sample was stored again
148 in the formaldehyde solution. Determinations of zoea I larvae of *H. sanguineus* were carried
149 out with the key of Hwang et al. (1993) and in comparison with larvae of the same species
150 obtained from ovigerous females collected in the intertidal. North European waters host three
151 species with similar larval morphology, *Eriocheir sinensis*, *Hemigrapsus sanguineus* and
152 *Hemigrapsus takanoi*. *E. sinensis* was distinguished from *Hemigrapsus* spp. by differences in
153 the morphology and setation of the telson (Ingle 1992, Hwang et al. 1993). Out of the
154 *Hemigrapsus* species, only *H. sanguineus* is abundant on Helgoland while *H. takanoi* is rare
155 (Jungblut et al. 2017).

156 References

157 Hwang, S.G. et al. 1993. Complete larval development of *Hemigrapsus sanguineus* (Decapoda,
158 Brachyura, Grapsidae) reared in laboratory. – Korean J. Syst. Zool. 9:69–86.

159 Ingle, R. 1992. Larval stages of Northeastern Atlantic crabs: an illustrated key. – Chapman and
160 Hall.

161 Jungblut, S. et al. 2017. Population development of the invasive crab *Hemigrapsus sanguineus*
162 (De Haan, 1853) and its potential native competitor *Carcinus maenas* (Linnaeus, 1758) at
163 Helgoland (North Sea) between 2009 and 2014. – Aquat. Inv. 12:85–96.

164

165 **Appendix 5. Methods: Platforms used to compute WLR_{15}**

166 Predictions for N. America were based solely on temperature data available in Emodnet data
167 portal (Emodnet platforms 44007, 44027, 44032, 44033, 44034, argm1, cfwm1, psbm1); data
168 of Seatemperatures.org did not provide information in the spatial resolution needed to evaluate
169 if WLR_{15} would match the distribution limit of *H. sanguineus*. Data from Emodnet enabled to
170 explore year-to-year variations in predictions of WLR_{15} (years 2016-2018). In some time
171 series, TW_{15} is not made of a series of consecutive days with T >15°C; instead the window was

fragmented because of temperatures dropping down (usually to ~13°C) for a few days. We minimized such effects by computing WLR_{15} after the time series was smoothed with a 3-point centered moving average. There were, however, two cases where the window was still fragmented. In those cases (highlighted in the section of Results) WLR_{15} is overestimated because temperature dropped below 15°C for long periods.

In Figure 5, platform names and coordinates are given in (Lat, Long) are found in Emodent. They are ordered from NE to SW as follows: 1: psbm1 (44.905, -66.983); 2:cfwm1 (44.657, -67.205); 3: 44027 (44.283, -67.3); 4: 44034 (44.103,-68.112); 5: atgm1 (44.392, -68,204); 6: 44033 (44.055, -68,996); 7: 44032 (43.715, -69,355); 8: 44007 (43.525,-70.141). For reference: some platforms correspond to cities as follows: 1: Eastport; 5: Bar Harbor; 8: off Portland.

Predictions of WLR_{15} for Europe were based on temperature data retrieved from both Seatemperature.org and Emodnet. Seatemperatures.org provided more detailed spatial information along the North Sea than Emodnet. To obtain daily values of temperature from Seatemperatures.org we smoothed the time series of monthly mean temperatures with a LOESS; the fitted model was then used for interpolation of daily values. Data from Emodnet (platforms Heimdal, Helgoland WR-North, Frderikshavn) were then used as a check of WLR_{15} at three available reference sites. Such data were treated as in the case of the times series for N. America (i.e. smoothed through a 3-point centered moving average window). In this case, we were also able to evaluate the effect of year-to-year variations on WLR_{15}, in particular the role of the heatwave experienced in summer 2018.

In Figure 6, In (a): Site names as follows: 1:Alesund, 2:Bergen, 3:Haugesund, 4:Stavanger, 5:Kristiansund, 6:Arendal, 7:Sandefjord, 8:Hirsthals, 9:Frederikhavn, 10:Thyborøn, 11:Hvide Sande, 12:Nordby, 13:Sylt, 14:Helgoland, 15:Lowestoft, 16:Hull, 17:New Castle, 18:Berwick, 19:Aberdeen. In (b): Platform names and coordinates (Lat, Long) are found in Emodnet. They are ordered from NE to SW as follows: 2: Heimdal (59.5742, 2.2273), 9: Frederikshvn: (57.4427, 10.5210). 14: Helgoland (54.1789, 7.8853). Platform numbers are chosen so as to coincide with site numbers.

Appendix 6. Data for validation of TTM

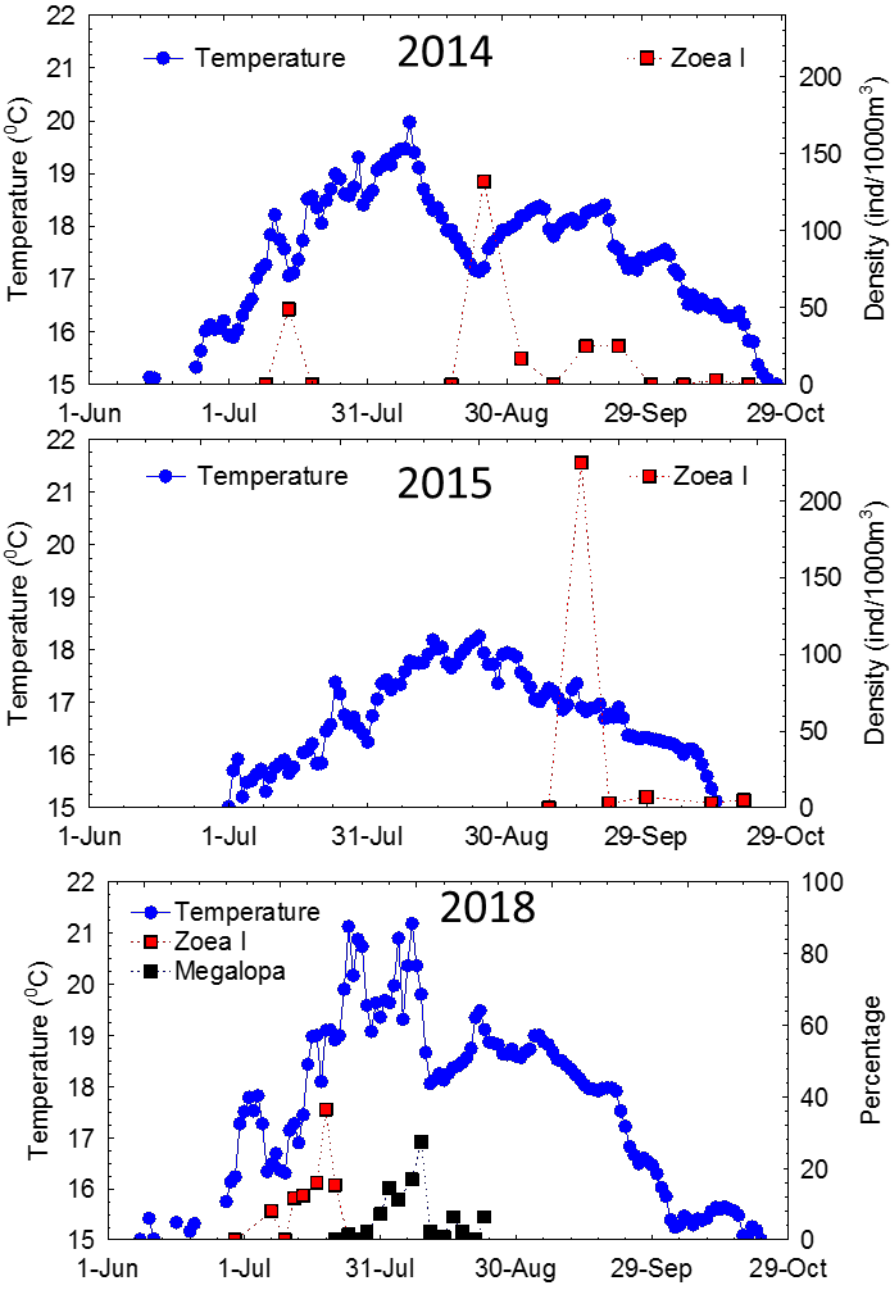

Figure A3. *Hemigrapsus sanguineus*. Daily temperatures (source COSYNA) and temporal distribution of larvae. Zoea I larvae were collected 1 to 3 times a week with a plankton net towed at the Helgoland roads in the summers of 2014, 2015, and 2018. Megalopae were collected using floating traps deployed in the rocky intertidal of Helgoland in 2018 (no collections of megalopae were made in 2014 and 2015). Data for 2014 and 2015 are presented as densities, but data of 2018 are presented as percentage to standardise abundance of zoea I and megalopa to the same scale.

Appendix 7. Results: laboratory experiments: Model selection and model fit

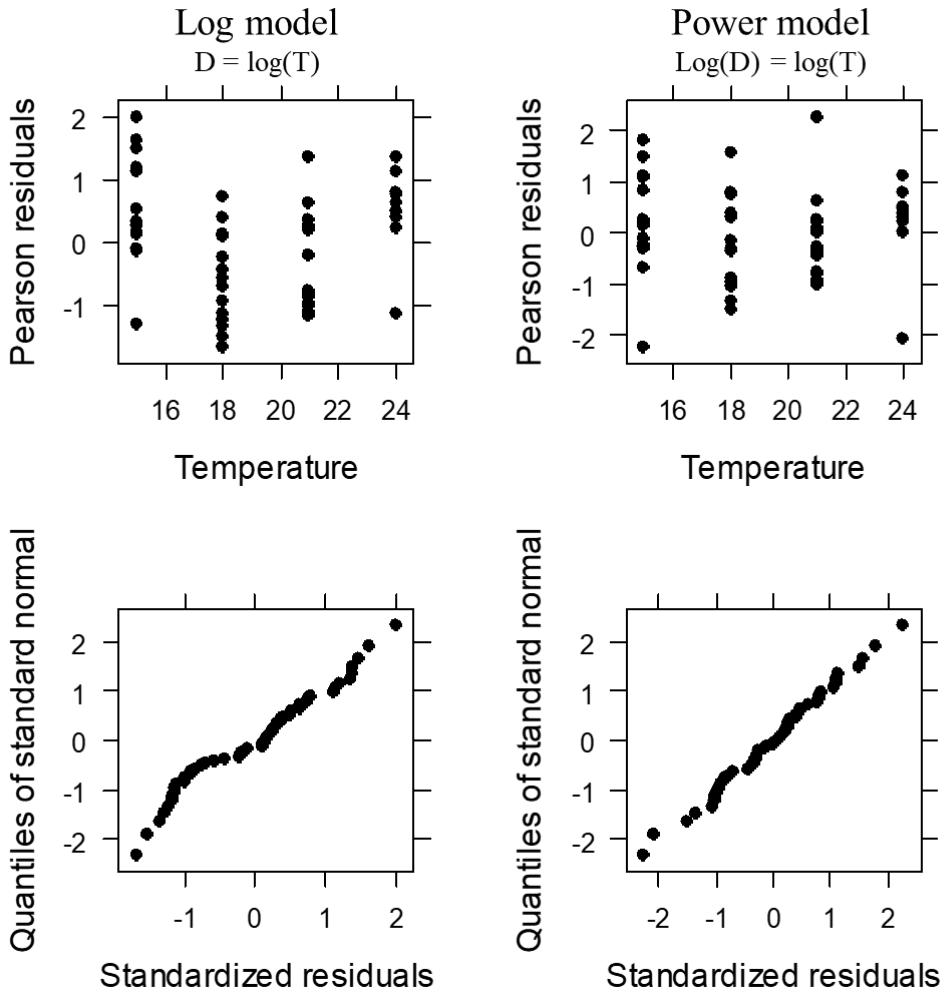

Figure A4. *Hemigrapsus sanguineus*. Selection of best model explaining relationships between temperature (T in °C) and duration of development to megalopa (D in days). The two candidate models are shown at the top of the panels. Top panels: residuals plotted against temperature: for the log-model (top left), there is a strong indication of nonlinearity (residuals at 18 °C are lower than at other temperatures). This trend is not present in the power model (top right). Bottom panels: quantile distributions: the log-model (bottom left) shows a bump, deviating from the straight line. The power model (bottom right) shows an almost straight pattern indicating that residuals are nearly normally distributed.

Table A1. *Hemigrapsus sanguineus*. Parameter estimates of the fitted model predicting the effect of temperature (T) on duration of development from hatching to megalopa (D). Abbreviations: SE: Standard error, SD: standard deviation. The model fitted corresponds to log(D) ~ log(T) with random intercepts and slopes. There were 52 observations spread collected in larvae from four females. The fixed effects correspond to the parameters, of the power model $D= aT^b$ with Intercept = log(a) and slope = b. Random effects give the standard deviation of the intercept and slope as driven by variation depending on the female of origin.

	Estimate	**SE**
Fixed effects		
Intercept	9.8934	0.5945
Slope	-2.2244	0.1943
Random effects	**Estimate**	**SD**
Intercept	1.34631	1.16031
Slope	0.14345	0.37875
Residual	0.00225	0.04743

Appendix 8. Results: field data:

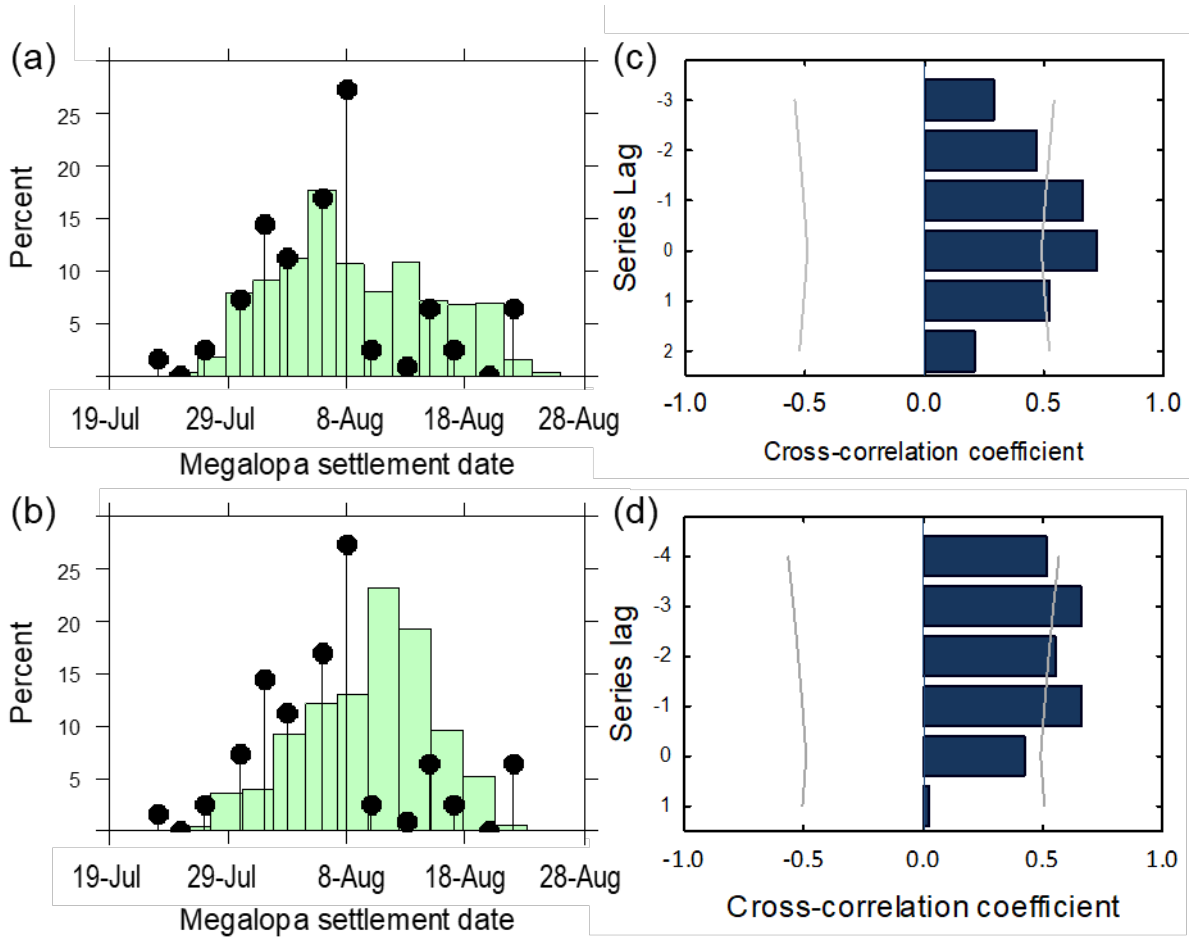

Figure A5. *Hemigrapsus sanguineus.* Same as Fig. 4, but adding the cross-correlations (c and d) between observations and predictions for the time series of observed abundance (a and b); grey lines delimit the 95% confidence interval obtained from white noise; correlations outside the limits are considered significant.

Appendix 9. Simulating effects of WLR using a metapopulation model

We simulated the effect of temperature on species distribution through the window of larval release, based on the model presented in Figure A1a. We did three separate model runs, based on the temperature data available for three years (2016-2018), but keeping the temperature constant in each model run. Each model run was as in the previous model: the first 40 years were run for 40 generations with a highly connected matrix; then, the second 40 years were run with assuming that populations were isolated. We did not allow for variation in temperature among years because we did not have a time series of temperature of 40 years of duration. Instead, the model simulations explore the effect of the breadth WLR on the range limits.

The model based on a coastline (as in Fig. A1a), with 16 populations (as defined in Fig. 5, main manuscript). A coastal cell is assigned to each local population, but now each cell is characterised by a value of WLR and duration of larval development assigned from a nearby temperature platform (Table A2). In the new version, the larval survival is driven by the effect of temperature on duration of development, existing in each coastal cell. In addition, the number of larvae settling in a site is now proportional to WLR assigned for a coastal cell (for larvae retention) and to the number of days when the WLR of adjacent population are simultaneously open (for larvae dispersing to and from adjacent sites).

Table A2. Window of larval release (WLR) and proportional larval survival (L_S) calculated for larvae of each local population. Site codes and crab numbers corresponds to those given in Lord & Williams (2017). P = platform number providing data for WLR and Ls; platform number are as follows: 1: psbm1; 2:cfwm1; 4: 44034; 5: atgm1; 6: 44033; 7: 44032; 8: 44007.

Site		# Crabs	P	WLR			Ls (x 10^{-3})		
#	Code	(ind m^{-2})		2016	2017	2018	2016	2017	2018
1	OD	9	8	52	18	55	2.98	1.89	4.48
2	KT	31	8	52	18	55	2.98	1.89	4.48
3	YK	11	8	52	18	55	2.98	1.89	4.48
4	BP	5	8	52	18	55	2.98	1.89	4.48
5	CE	9	8	52	18	55	2.98	1.89	4.48
6	LE	11	7	34	13	45	1.46	0.75	2.25
7	PP	3	7	34	13	45	1.46	0.75	2.25
8	MP	1	6	35	7	36	1.10	1.23	1.52
9	OH	1	6	36	1	27	1.10	1.23	1.52
10	MO	0	5 (4)	3	0	2	0.70	0.00	1.16
11	ST	0	5 (4)	3	0	2	0.70	0.00	1.16
12	BH	0	5 (4)	3	0	2	0.70	0.00	1.16
13	LM	0	5	6	0	4	0.70	0.00	1.16
14	PM	1	5	6	0	4	0.70	0.00	1.16
15	JP	0	2	0	0	0	0.00	0.00	0.00
16	CB	0	1	0	0	0	0.00	0.00	0.00

297 Because only few platforms provide data of a region with 16 local populations, we assigned
298 the nearest platform to each local population (Table A2). At the limit of the distribution range
299 there are three out of four platforms giving WLR = 0; For simplicity, we calculated D as the
300 average of the two nearest platforms (the one with WLR>0 and the nearest platform giving
301 WLR=0), instead of averaging all platforms. This approach was carried out in order to avoid
302 forcing the limit of distribution to coincide with that observed in the data by Lord & Williams
303 (2017); it should instead overestimate the degree of northward expansion.

304 Larval survival rates (Table A2) were calculated as:

305 $$Ls = e^{-\mu \cdot D} \quad (7)$$

306 where μ is the daily instantaneous larval mortality rate (= 0.18, based on Rumrill 1990, Wilson
307 et al. 2014) and D is the duration of larval development, estimated from the TTM models for
308 each coastal cell separately. For larvae retained within a site j, L_Sjj was calculated as in equation
309 (7). For those migrating to an adjacent cell, survival is calculated following eq. (4) , LSij=
310 $(L_Sjj)^{0.5} \cdot (L_Sii)^{0.5}$, i.e. as the product of survival rates expected for each cell given that larvae
311 spend half the time within each cell as assuming independence.

312 The model is set with the following parameter values: Adult fecundity (*f=1000*), dispersal
313 matrix as in Fig. A1a, juvenile survival defined by a Beverton-Holt function, with parameters
314 ($\alpha = 0.3$, $\beta = 0.01$). The number of juveniles was calculated as

315 $$J_i = 0.3/(1+0.01\ S_i), \quad (8)$$

316 where *Si* is the number of settlers (following equation 3 in Section 1). In the first step, the
317 model is run for 40 time-steps to produce Figure 7a, with $N_0 = 10$ for each local population. In
318 the second step, the connectivity matrix has only retention coefficients (= 1) and the model is
319 run for additional 40 times, using the abundance patterns obtained at the 40th iteration in the
320 first step.

321 The parameter values result in recruitment success with WLR ≥ 1 and D ≤ 30 assuming full
322 retention (diagonal terms of the dispersal matrix = *dii=1* and all off diagonal terms, *dij =0*).
323 One can predict population changes at low density (density-dependence ignored) by the ratio
324 $N(t+1)/N(t) = R = f \cdot WLR \cdot L_S \cdot \alpha\ dii$: the population will go extinguished if R <1. Under the
325 parameter combination used here we get R= 1.35 at WLR=1. Once the density-dependent effect
326 is considered (through $\beta = 0.01$) the populations stabilises at a maximum of ~30 individuals,
327 coinciding with the maximum average densities reported by Lord & Williams (2017).

328 The model is run through matrix algebra as in the original case. Using the inner product ($\times$)
329 and the Hadamard product ($*$), we obtain a new spatial projection matrix described in eq. 5 that
330 is now computed as:

331 $$M1 = (D \times F) * L \text{ (9a)}$$

332 $$SPM = M1 * W \text{ (9b)},$$

333 where M_1 is the product of the fecundity (F), larval dispersal (D) and larval survival (L). The
334 matrix W introduces the window for larval release, with elements *wii* and *wij*. This matrix has,
335 as diagonal elements (*wii*), the values of WLR assigned to each site depending on the distance
336 to the nearest temperature platform (see Table A2). The subdiagonals (*wij*) correspond to the
337 number of days in the year when the windows of adjacent populations are open at the same
338 time. In this way, we obtain the number of settlers in any population *i* (*Si = sii+sij*; *Si* was
339 defined in eq. 8) as proportional to WLR. For larvae returning to the same population we obtain
340 *sii = dii·f·Lii · wi* (diagonal) and for those arriving from other populations we obtain *sij = dij·f*
341 *·Lij· wi* (subdiagonal and supradiagonal) but other terms are *sij = 0* (larvae only disperse to
342 adjacent sites). The number of larvae arriving to each local habitat is then computed as the
343 inner product of SPM and the number of individuals per population (equation 6, Section 1),
344 with 16 local populations. As in the model of section1, S is used to compute a vector of juvenile
345 survival probabilities *J*, so that *N(t+1) = S · J* (element-by-element multiplication).

346