Selective Bird Predation on the Peppered Moth

Electronic supplementary materials ESM3

L M Cook, B S Grant, I J Saccheri, J Mallet

Statistical analysis details for results in table 2 of the main paper

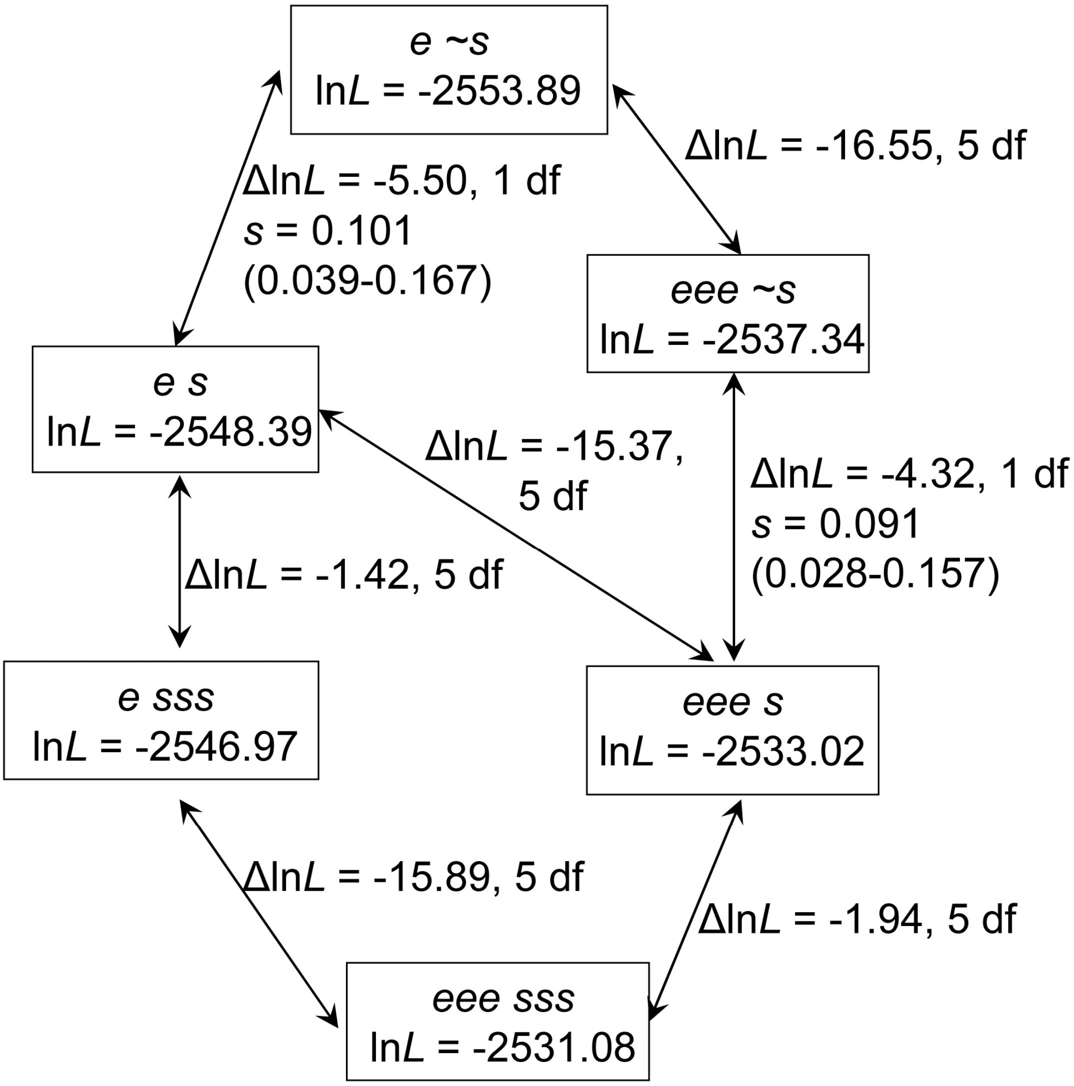
Suppose that, in each year i the probability of not being eaten is $(1-e_i)$ for *typica*, and $(1-e_i)(1-s_i)$ for *carbonaria*, where $0 \leq e_i < 1$, and $s_i < 1$. Then e_i measures the non-selective probability of being eaten (independently of phenotype), s_i measures the additional or 'selective' probability of being eaten, measured as a selection coefficient against *carbonaria*; $(1-s_i)$ can thus be interpreted as the relative fitness of *carbonaria* in this experiment.

The expected probabilities in each year are then as follows:

Eaten		Not eaten	
typica	*carbonaria*	*typica*	*carbonaria*
e_i	$1-(1-e_i)(1-s_i)$	$(1-e_i)$	$(1-e_i)(1-s_i)$

Maximum likelihood parameter estimates, log likelihoods ($\log_e L$) for various different models varying in complexity were found, and support limits for s, were calculated. Twice the log likelihood difference (or deviance, $G = 2\Delta\log_e L$) between nested models has an approximately chi-squared distribution under the simpler model, and was used to test whether a model of higher complexity gave a better fit, yielding a multidimensional contingency table, analysed as a log-linear model [1]. Two unit log likelihood-based support limits are asymptotically equivalent to 95% confidence limits [2,3], and were obtained conditional on remaximizing likelihoods over all other estimated parameters. All calculations and maximizations were carried out using spreadsheet software. Although the models are parameterized using selection and fitness, the analysis is exactly equivalent to that used in generalized linear models.

Some individuals are always eaten, so e is always a non-zero parameter. The simplest model to explain the data, with 1 parameter estimated, is then "e ~s", giving an estimate of e but with $s = 0$; i.e. *carbonaria* is assumed to have no extra selective mortality. The most complex model giving a perfect fit is "*eee sss*," with 12 estimated parameters (the maximum) allowing variable e_i and variable s_i across years, *i*. The different complexities of models form a hierarchy which allow likelihood ratio tests of hypotheses (see the somewhat similar diagram of three-way frequency analysis in ref. [1]).



The results show:

i) strong evidence of heterogeneity in *e* across years, with $P < 0.001$ in both possible tests, but

ii) no evidence of significant variation in *s* across years (P > 0.5 in both possible tests), and

iii) strong evidence of selection in tests of ~*s* vs. *s*, with $P = 0.003$ (or $P = 0.001$ under the simpler assumption that *e* is constant, i.e. based on simple totals over all years, *e ~s vs. e s*), and also that

iv) the maximum likelihood daily selective coefficient, with 2-unit log-likelihood confidence limits, under the most conservative assumptions (*eee s*), is $s = 0.091$ (0.028-0.157). (All other parameters were remaximized during estimation of 2-unit limits). This gives daily relative survival estimates for melanics of 91% (84%-97%) of the typical forms. Assuming that e is constant over years, a simpler estimate is $s = 0.101$. This latter estimate is less conservative, but the two values hardly differ.

References

1 Sokal, R.R. & Rohlf, F.J. 1994 *Biometry. Third Edition*. San Francisco: Freeman.

2 Edwards, A.W.F. 1992 *Likelihood. Expanded Edition*. Baltimore, Maryland: Johns Hopkins University Press.

3 Pawitan, Y. 2001 *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford: Oxford University Press.