

Chung, N., Pienaar, J. and Greeff, J. M. 2019. Evolutionary stable sex ratios with non-facultative male-eggs first sex allocation in fig wasps. – Oikos doi: 10.1111/oik.06068

Appendix 1

Deriving parameters for mortality distribution

The distribution of egg deaths per patch has two parameters: p (probability) and σ^2 (variance).

These two parameters can be written in terms of μ (mean) and r (dispersion):

$$p = \frac{r}{r + \mu} \quad (\text{A1})$$

$$\sigma^2 = \mu + \frac{\mu^2}{r} \quad (\text{A2})$$

Equation A1 and A2 can be reworked into Eq. A3 and A4 respectively:

$$p = \frac{\mu}{\sigma^2} \quad (\text{A3})$$

$$r = \frac{\mu^2}{\sigma^2 - \mu} \quad (\text{A4})$$

The p and r parameters are calculated from σ^2 and μ . In turn, σ^2 and μ can be further calculated from two input parameters: D_m (egg death mean) and D_v (egg death dispersion). D_m and D_v are either specified or estimated (estimates were originally calculated from single foundress figs). The μ term is the sum of eggs laid per fig multiplied by D_m divided by F_e :

$$\mu = \sum_{i=1}^{k_{max}} l_i \left(\frac{D_m}{F_e} \right) \quad (\text{A5})$$

The σ^2 term is the sum of eggs laid per fig multiplied by D_v divided by F_e :

$$\sigma^2 = \sum_{i=1}^{k_{max}} l_i \left(\frac{D_v}{F_e} \right) \quad (\text{A6})$$

It can be seen that if total eggs laid is equal to egg load λ , $\sum l_i = F_e$ (e.g. a single foundress lays an average number of eggs), then $\sum l_i$ and F_e will cancel each other out and consequently: $\mu = D_m$ (Eq. A5) and $\sigma^2 = D_v$ (Eq. A6). For a single foundress, if there is no oviposition constraint (i.e. $e_l < o$), the foundress will be able to lay her full egg load $l_l = e_l$. Since e is sampled from F_e (which is also the denominator in Eq. A5 and A6), D_m and D_v are calculated relative to the average eggs laid by a single foundress. However, when the number of available oviposition sites is less than egg load ($O_a < F_e$), as can be the case in species with high levels of oviposition constraint, then the denominator in Eq. A5 and A6 will be O_a instead of F_e . In these cases, F_e is no longer a good proxy for average eggs laid per single foundress since egg loads will be constrained by O_a . Incorrectly using F_e instead of O_a would lead to lower than expected deaths relative to averages calculated in the species parameter estimation methods.

Details of the resampling process

This resampling process is illustrated by the following example. Foundress A laid 6 male and 40 female eggs. Foundress B laid 4* male and 42 female eggs. This gives a total of 92 eggs in the patch. There is a randomly sampled death rate of $d = 24$ in the patch. d_M then randomly samples a multinomial distribution with four offspring classes based on the probability weighting of the classes ($f_{Am} = 6/92, f_{Af} = 40/92, f_{Bm} = 4/92, f_{Bf} = 42/92$). If for instance $d_M = [2, 10, 5*, 7]$ then death for foundress B male eggs would be -1 . This would prompt a resampling event. If the next randomly sampled $d_M = [4, 10, 1, 9]$ this would not prompt a resampling event because none of the d_M classes produces a negative offspring egg classes.

At low to moderate mortality or relatively equal sized offspring class, the resampling procedure produces proportional deaths in the offspring classes. However, when death rates are high or offspring classes are drastically different, then the resampling process has a tendency to produce lower than proportional deaths for the smaller class. Since male numbers are typically much lower than female numbers, there is a greater tendency for negative males to prompt a resampling event. This tends to slightly bias survival towards males in the simulations.

Parameter estimation methods

The parameter that specifies the λ of the distribution of usable ovules (O_a) was estimated iteratively by setting the largest observed offspring count (from either single- or multi-foundress offspring

counts) as the 95th percentile of a Poisson distribution. A range of possible Poisson means were simulated a default of 10,000 times each until the one that produced a value that matched the 95th percentile was found.

Although the distribution of available ovules (O_a) is critical to the model (since oviposition constraint is determined by this parameter), estimates proved difficult to verify (Table A1). Even if figs had been dissected and all ovules counted, this would still not be a direct estimation of available ovules since some outer ovules may be inaccessible to the foundress or non-pollinator wasp could occupy some portion of the ovules. The O_a estimate for *F. lutea* (pollinated by *A. heterandromorphum*) was based not on the largest observation (total offspring = 514) but the second largest (total offspring = 296), due to the fact that the former was an extreme outlier. Since the sample size of *F. sycomorus* (pollinated by *C. arabicus*) was very small, O_a was estimated using the maximum observed offspring count (total offspring = 606) for *Ceratosolen galili* (Pentz and Greeff, pers. comm.), an internally ovipositing ‘cuckoo’ wasp that is distantly related to (Kerdelhué et al. 1999) and similar in size to the true pollinator *C. arabicus*.

Table A1. Available ovules (O_a) parameter estimates. The largest observed count (total offspring = 514) for *F. lutea* was excluded due to being an extreme outlier. Since the sample size for *F. sycomorus* was very small ($n = 9$), the largest observation for *C. galili* (Pentz and Greeff, pers. comm.) was used to estimate O_a instead.

Tree species	Pollinator wasp	Observed	Estimated
		Max	O_a
<i>F. abutilifolia</i>	<i>E. comptoni</i>	294	268
<i>F. burkei</i>	<i>E. stuckenbergii</i>	120	103
<i>F. glumosa</i>	<i>E. glumosae</i>	87	73
<i>F. ingens</i>	<i>P. soraria</i>	129	112
<i>F. lutea</i>	<i>A. heterandromorphum</i>	296*	269
<i>F. salicifolia</i>	<i>P. awekei</i>	85	71
<i>F. sycomorus</i>	<i>C. arabicus</i>	606**	565

The two parameters for the foundress distribution (F_s and F_p) which specify a zero-truncated negative binomial distribution were estimated via simulation. Initially, a wide range (e.g. 0.1 to 0.9) was set for the starting and end points for both parameters for which simulations would be performed. The interval between points at this stage was set to 0.1. These simulated candidate distributions were then compared to the observed species distribution by calculating the residual sum of squares (RSS) between them (Eq. A7). The combination of candidate variables that produced the lowest RSS were then simulated a second time with a narrower range (e.g. 0.25 to 0.35) using a smaller interval (0.01) until fine-grained estimates for the two parameters were found.

The candidate distribution yielding the lowest RSS point was the estimated distribution. Each simulation was performed a default of 10 000 times.

$$RSS = \sum_{i=1}^{k_{max}} (obs_i - exp_i)^2 \tag{A7}$$

To test the fit of the estimated distribution to the observed, the estimated distribution was simulated a default of 1 000 000 times and scaled proportionally to produce the same sum of observed foundresses. Then the two distributions were binned into as many bins as there were simulated or observed foundress counts. A Pearson's goodness-of-fit test was performed where p-values were simulated by Monte Carlo simulation a default of 1 000 000 times. In the case where bins had zero observations, the bin was set to an expected value of 1×10^{-10} which effectively makes it zero but still permitted the calculation of a test statistic. Foundress parameter estimates are shown in Table A2. Figure A1 shows an example of estimated and observed foundress distributions for *F. abutilifolia* (pollinated by *E. comptoni*).

Table A2. Foundress (F_s and F_p) parameter estimates. N: total number of counts used for parameter estimation. Arth. mean: arithmetic mean. Har. mean: harmonic mean. % 1F: fraction of single foundress figs.

	Observed				Estimated parameters				Estimated	
	N	Arth. mean	Har. mean	% 1F	F_s	F_p	p	Arth. mean	Har. mean	% 1F
<i>F. abutilifolia</i>	347	2.63	1.60	0.43	0.54	0.32	0.99	2.50	1.59	0.43
<i>F. burkei</i>	301	1.33	1.15	0.76	0.68	0.72	1.00	1.32	1.15	0.76
<i>F. glumosa</i>	246	1.60	1.21	0.70	0.33	0.58	0.93	1.45	1.20	0.70
<i>F. ingens</i>	54	1.35	1.13	0.80	0.34	0.72	1.00	1.25	1.12	0.81
<i>F. lutea</i>	97	1.94	1.47	0.45	0.80	0.42	0.43	2.21	1.51	0.46
<i>F. salicifolia</i>	628	1.79	1.38	0.54	0.84	0.52	0.72	1.84	1.37	0.55
<i>F. sycomorpus</i>	185	1.44	1.21	0.69	0.68	0.64	0.94	1.46	1.21	0.69

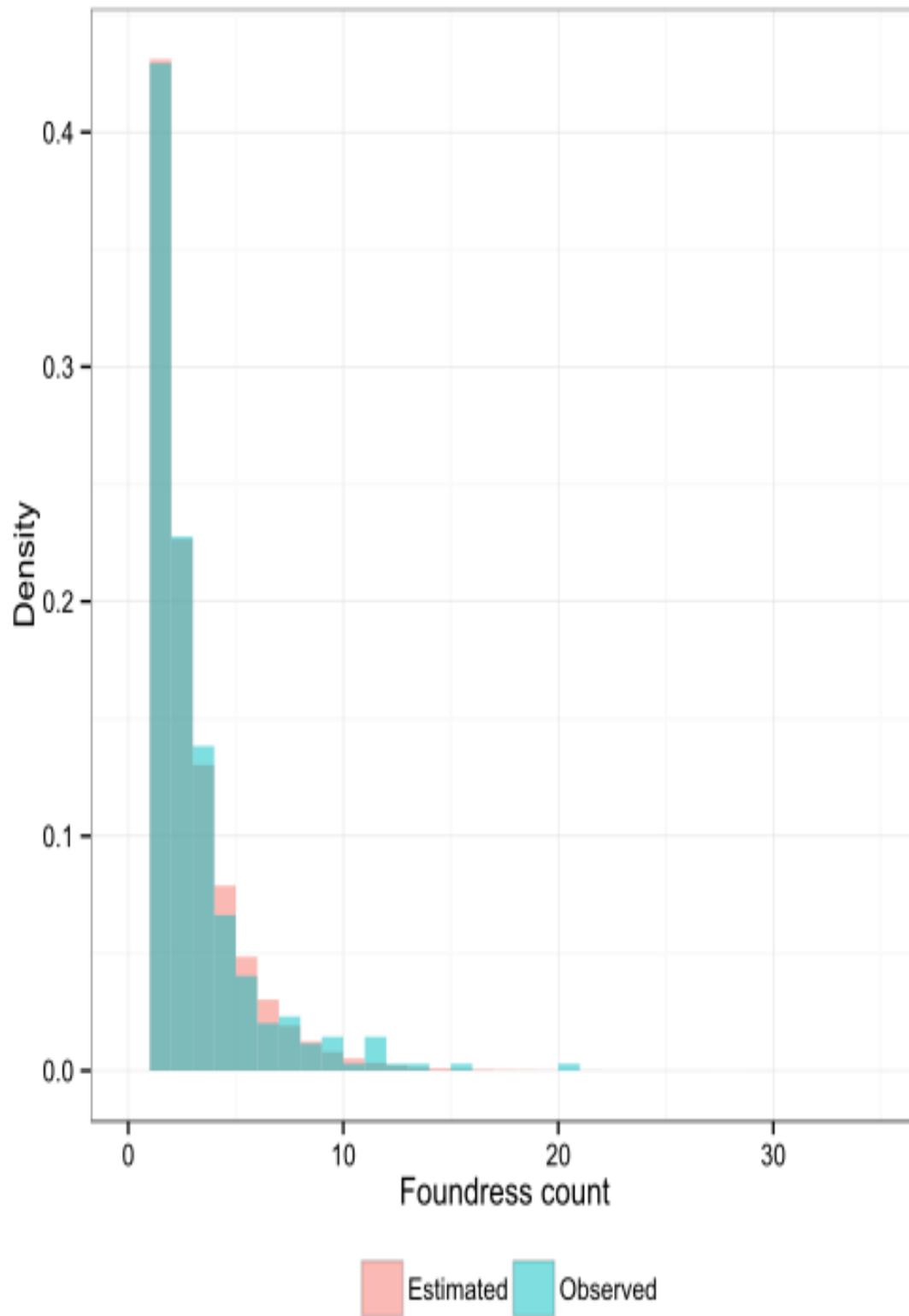


Figure A1. Estimated and observed distribution of *E. comptoni* (pollinator of *F. abutilifolia*) foundresses counts per fig.

Egg load (F_e), death mean (D_m) and death variance (D_v) parameters were estimated at the same time. The process is similar to the MEN model algorithm described above: we assume egg load is modelled by a Poisson distribution and that egg mortality is modelled by a negative binomial distribution. It should be noted that the distribution of egg loads is not the same as the distribution of eggs laid since a shortage of ovules and/or competition with other foundresses for sites may result in not all eggs being laid. Each fig has a number of usable ovules (O_a) randomly sampled from the distribution previously estimated above. The number of usable ovules is not a direct representation of the distribution of ovules in a fig. Outer ovules may be inaccessible to foundresses or be occupied by non-pollinating fig wasps. The number of ovules suitable for oviposition by the pollinator foundress will be lower than the total number of ovules in a fig.

Given one mated foundress per fig, we simulated a range of candidate distributions generated by a specified range of candidate variables for F_e , D_m and D_v values. The candidate distributions were then compared to the observed distribution in single foundress counts. Estimation then proceeded in the same way as for F_s and F_p except that the observed and estimated distributions were binned into 10 equally sized bins before performing a Pearson's goodness-of-fit test.

Death rate and death variance are not independent parameters since they will have to vary proportionally with the range of simulated egg loads; the difference between mean egg load and mean mortality has to equal the observed offspring mean. After the eggs are laid, the egg mortality process occurs which reduces the number of eggs. This results in a longer left tail distribution of eggs survived (i.e. offspring). Because offspring numbers are bounded at zero (clutch size cannot be negative), the distribution will become zero-inflated since all cases where death exceeds the number of eggs defaults to a clutch size of zero. Since no count data was collected for total brood mortality, we did not include the zero-count class in the RSS calculation. As long as death rate is low to moderate, the absence of the zero-count class comparisons does not bias the parameter estimation method too much.

Depending on the level of constraint or competition, overestimating the egg load parameter (F_e) will not significantly alter the model predictions since what matters more is the eggs actually laid (which is a function of O_a when constrained). The variance of the egg distribution is less than the variance of the offspring distribution since mortality results in overdispersion.

Estimates for egg load (F_e) and egg mortality rate (D_m and D_v) were based on data that had potential outliers excluded (Table A3). The parameter estimation method is sensitive to outliers so to ensure accurate estimates, counts in which we suspected the counts may not be actual single foundress counts were excluded. The rule that was applied for exclusion was any count that was approximately more than double the mean (the potential outlier is not part of the mean calculation).

This process was judged entirely on the size of the clutches and was agnostic to their observed sex ratios. In *E. comptoni* (pollinator of *F. abutilifolia*), a count of 294 was excluded (the offspring mean without this count is 137). In *Elisabethiella stuckenbergi* (pollinator of *F. burkei*) for which we have samples from two trees, samples from one tree ($n = 10$) had very low offspring numbers and no males. The offspring here showed a bimodal distribution, whereas others species were mostly unimodal. But since it may be possible that such a high death rate is part of the natural variation of *E. stuckenbergi* (pollinator of *F. burkei*), we calculated two sets of parameter estimates: 1) '*F. burkei*-all' which contains samples from both trees and, 2) '*F. burkei*-rem' which excludes the samples from the tree with high deaths and no males. Since the simulation results were similar for both, we decided to use '*F. burkei*-all'. *Elisabethiella glumosae* (pollinator of *Ficus glumosa*) likewise showed a bimodal distribution but since the instances of high mortality was distributed across two trees, rather than all being confined to one, we decided to keep them. However, there was one offspring count of 65 that was removed (the offspring mean without this count is 17). For *A. heterandromorphum* (pollinator of *F. lutea*), one extreme offspring count of 514 was excluded (the offspring mean without this count is 200).

All p-values in Table A2 and Table A3 were generated by Monte Carlo simulation. Since some of the samples sizes were very small, these values may not be reliable statistics. Given this caveat, the estimated distributions are still the ones which had the lowest RSS values; no other distribution generated using the parameter estimation method would provide a better fit.

Table A3. Egg load (F_e) and mortality (D_m & D_v) parameter estimates. N: total number of counts used for parameter estimation.

Tree species	Pollinator wasp	Observed		Estimated parameters				Estimated
		N	Mean	F_e	D_m	D_v	p	Mean
<i>F. abutilifolia</i>	<i>E. comptoni</i>	18	144.33	191	56	3575	1.00	143.21
<i>F. burkei</i> -all	<i>E. stuckenbergii</i>	31	43.35	139	87	2275	0.40	42.59
<i>F. burkei</i> -rem		21	61.00	98	44	1400	1.00	60.01
<i>F. glumosa</i>	<i>E. glumosae</i>	18	17.11	47	36	175	0.12	17.02
<i>F. ingens</i>	<i>P. soraria</i>	29	19.10	128	102	275	0.45	19.11
<i>F. lutea</i>	<i>A. heterandromorphum</i>	66	199.79	265	63	3625	0.74	198.37
<i>F. salicifolia</i>	<i>P. awekei</i>	47	46.57	56	16	775	0.89	46.19
<i>F. sycomorus</i>	<i>C. arabicus</i>	9	110.56	324	231	5100	1.00	110.18