Appendix S1
Fitness reduction for uncooperative fig wasps through reduced offspring size: A third component of host sanctions

Jandér, K.C., Dafoe, A., Herre, E.A.

May 6, 2016

1 Additional Methods and Results

1.1 Length measurements correlate with wasp weight

To confirm that the measurements we use correlate with wasp weight we measured the length of the front femur, head and thorax to the nearest 0.01mm on 36 female wasps from *F. popenoei*, each having recently emerged from a different fig to ensure independence, then dried the wasps in 50C for 48 hours and weighed them to the nearest microgram. The length measurements for each individual were significantly correlated with wasp dry weight (Pearson correlations on log transformed data; weight vs thorax length: n = 36, r = 0.70, p = 2.0E-6; weight vs head length: n = 36, r = 0.77, p = 2.7E-8; weight vs femur length: n = 36, r = 0.81, p = 2.4E-9; Fig S1).
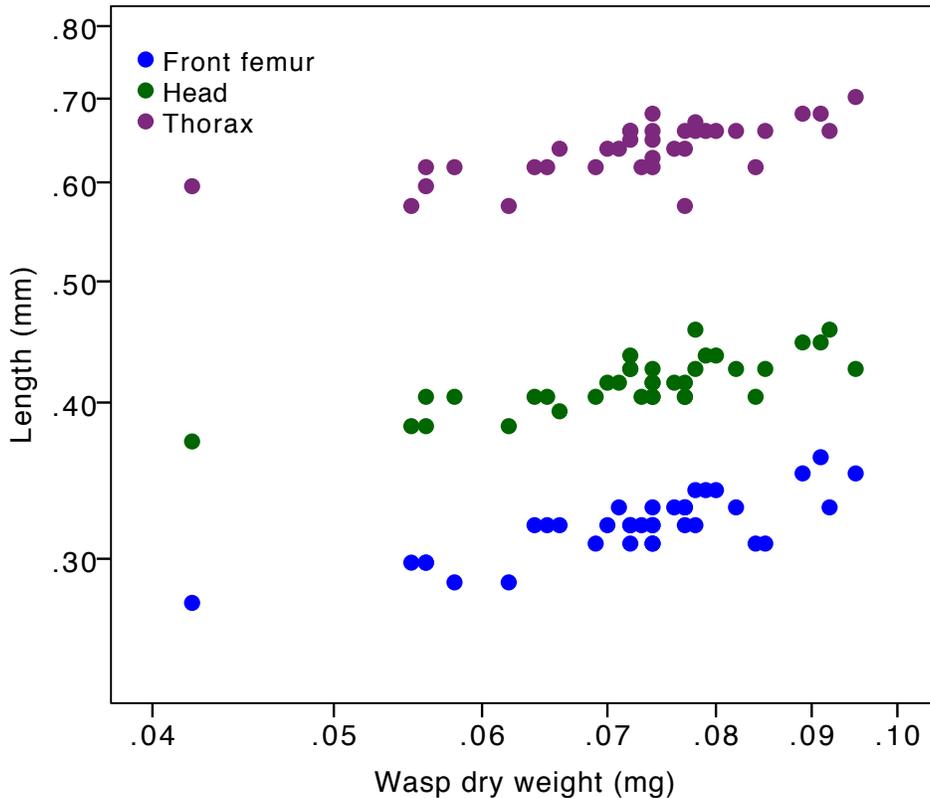

Figure S1. The dry weight of female pollinator wasps of *F. popenoei* was significantly correlated with front femur length, head length, and thorax length.

To be able to predict wasp head length from wasp dry weight outside the normal range for each species (a result of the P- treatment) we collected head length and dry weight for 20 to 37 wasp individuals from each of five species that vary in size (wasps from *F. citrifolia, F. nymphaeifolia, F. obtusifolia, F. popenoei* and *F. perforata*; the latter included because its tiny wasps well represent the range of size-reduced wasps for *F. citrifolia*). Together these species covered the entire range of wasp sizes produced in the experiments. There was a strong positive relationship between wasp dry weight and head length: (linear regression on logged data: Log head length (mm) = 0.40(Log dry weight (mg)) + 0.093; r^2 = 0.93, t_{134} = 42.43, p = 3.5E-79.

1.2 The P- treatment does not affect foundress size

Another assumption important for this study is that producing foundresses for the pollen-free treatment (opening the fig fruit before females have emerged in order to remove the pollen) does not affect foundress size. To test this we measured wasp front femur length, head length and thorax length on 38 P- and 36 P+ wasps from *F. popenoei*, each having emerged from a different fig fruit to ensure independence, then dried the wasps in 50C for 48 hours and measured their dry weight. There were no significant differences in the dry weight or length measurements between P- and P+ experimental wasps (t-tests; dry weight: t_{72} = 0.104, p = 0.92; front femur length t_{72} = 0.439, p = 0.66; head length: t_{72} = 1.33, p = 0.19; thorax length: t_{72} = 1.48, p = 0.14).
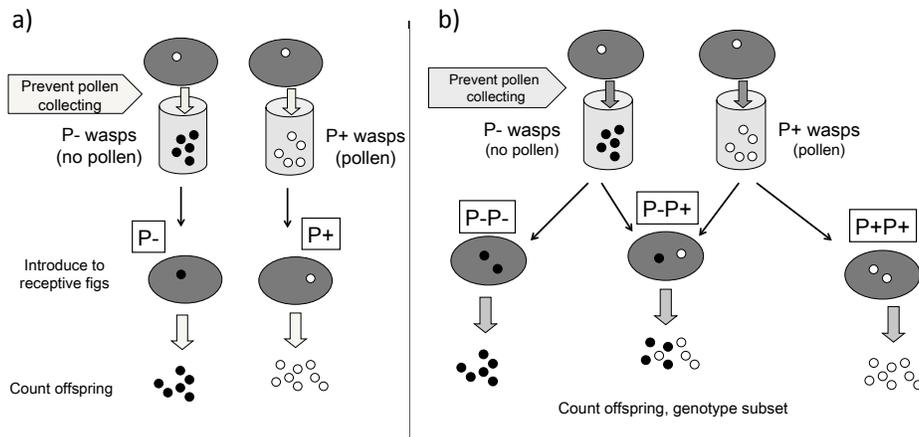
Figure S2. Diagram of the experimental setups A and B. Wasps were subjected to either the pollen-free (P-) or pollen-carrying (P+) treatment and transferred to vials. In experiment A, one foundress was introduced into each experimental fig. In experiment B, two foundresses were introduced into each experimental fig. The resulting offspring were collected when figs had matured.

Table S1. Predictions of offspring size based on the four hypotheses regarding the precision of sanctions. For example, under hypothesis 1A, the mean offspring size in P-P+ figs would, averaged over many figs, not differ between the P- and P+ lineages. Under the same hypothesis, P- offspring in P-P+ figs would on average be larger than those in P-P- figs. Likewise, P+ offspring in P-P+ figs would not differ in size from those in P+P+ figs.

Hypothesis	Within P-P+	P- in P-P+	P+ in P-P+
1A Fig level; 1P+ sufficient	$P- = P+$	$> P - P-$	$= P + P+$
1B Fig level; pollen-dependent	$P- = P+$	$> P - P-$	$< P + P+$
2A Flower & nearby level	$P- < P+$	$> P - P-$	$< P + P+$
2B Flower level only	$P- < P+$	$= P - P-$	$= P + P+$

Table S2. Wasp mean dry weights (mg) from P+/P- experiments on five different fig species; n denotes the number of figs. From each experimental fig 10 wasp offspring (where available) were dried and weighed, and the mean used in the t-tests.

Species; tree	P+ weight (sem)	n	P- weight (sem)	n	t (df)	p
F. cit; BCI17	0.0475 (0.00488)	10	0.0233 (0.00362)	3	-7.13 (11)	< 0.001
F. max; BS1	0.0723 (0.00212)	7	0.0651 (0.00163)	5	-2.5 (10)	0.031
F. nym; BN67	0.0878 (0.00284)	15	0.0638 (0.00193)	23	-7.24 (36)	< 0.001
F. nym; BS1	0.107 (0.00279)	15	0.0863 (0.00222)	11	-5.55 (24)	< 0.001
F. obt; BN64	0.145 (0.00400)	10	0.114 (0.0125)	4	-3.14 (12)	0.008
F. obt; MIL	0.122 (0.00378)	19	0.0984 (0.00229)	32	-5.60 (49)	< 0.001
F. pop; BV11	0.0468 (0.00154)	12	0.0462 (0.00113)	32	-0.29 (42)	0.77
F. pop; JG1	0.0409 (0.00530)	2	0.0375 (0.00146)	4	-0.87 (4)	0.43

Table S3. Details regarding the field-collected samples of wasp head length.

Species	crops (n)	foundresses (n)	emerging (n)
F. citrifolia	3	111	553
F. nymphaeifolia	2	121	605
F. obtusifolia	17	1178	5884
F. popenoei	3	179	894

1.3 Additional information on how we link the size of female offspring to fitness

1.3.1 Estimating $P(F|hl)$

We want to estimate the probability of a wasp of a given head length becoming a foundress. Denote this $P(F|hl)$, where hl denotes the head length of the wasp and F denotes that the wasp becomes a foundress. We have field-collected data on the head length for emerging wasps, allowing us to estimate the probability of an emerging wasp having a given head length $P(hl)$ (Fig. S3). We have data on the head length for wasps that become foundresses, allowing us to estimate the probability of a foundress having a given head length $P(hl|F)$ (Fig. S3). We have data on the baseline probability of a wasp becoming a foundress for each species $P(F)$. From this we can use Bayes' rule to calculate the desired quantity:

$$P(F|hl) = \frac{P(hl|F)P(F)}{P(hl)}$$

To implement this, we first estimate the probability densities $P(hl|F)$ and $P(hl)$ using a kernel density estimator. Because the data was measured in intervals of 0.01, we would prefer a density estimator appropriate to this binning; we do not want our density estimator to fit the absence of data between measurement intervals. We considered using the estimator described by Blower and Kelsall (2002). This density estimator is implemented in the **R** package **bda** as the "smkde" estimator type in the `bde` function, however the density estimate seemed to depend excessively on the random seed that we used. Instead we used the generic function `density` in **R**, with a bandwidth of 0.01 and the Gaussian kernel. We tried a variety of other bandwidths and bandwidth selection methods; the results appeared robust to them though the densities became much less smooth with smaller bandwidths in a manner suggestive of them fitting noise more than signal.

After estimating the densities we then calculated $P(F|hl)$.

To produce confidence intervals for $P(F|hl)$ we employed a nonparametric bootstrap: we repeatedly resampled from the data to produce bootstrapped samples. Specifically we resampled a foundress sample from the original foundress sample, of the same size as the original foundress sample; we resampled an emerging sample from the original emerging sample, of the same size as the original emerging sample. We then reestimated the densi-
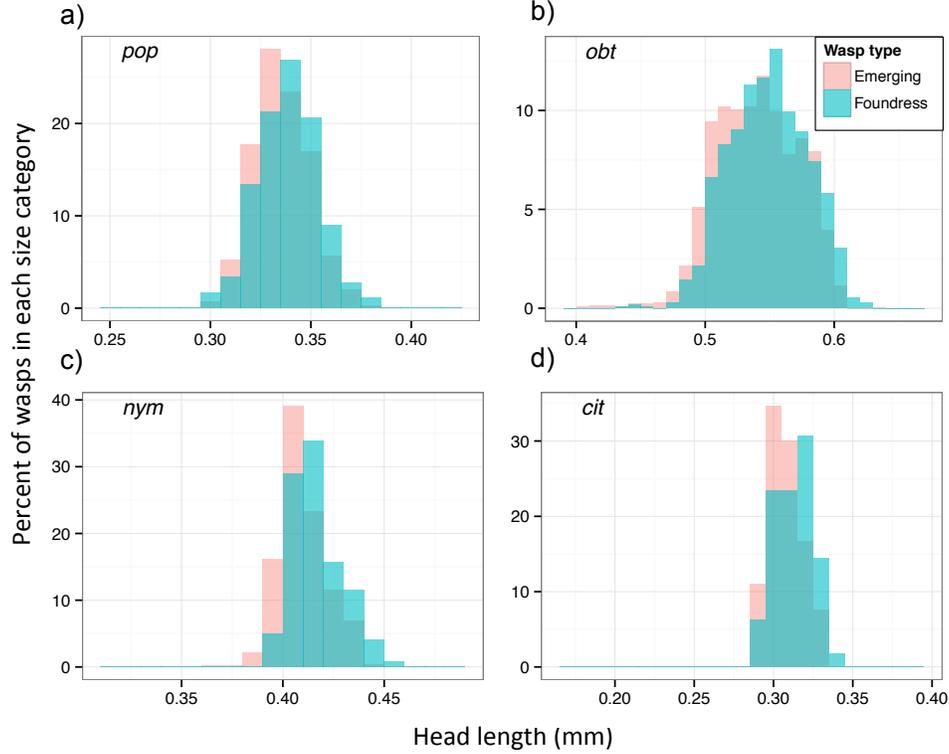
Figure S3. Distribution of head lengths of field-collected foundresses and emerging pollinator wasps associated with four fig species.

ties and $P(F|hl)$ for each bootstrapped sample. We then plotted the lower 2.5% and upper 97.5% quantile of the estimates of $P(F|hl)$ for each value of head length (hl) (Fig. 3).

1.3.2 Calculating average $P(F|hl)$ for P+ and P- offspring

We want to estimate the relative probability of becoming a foundress for P-offspring compared to P+ offspring. Denote hl_i^- as the counterfactual head length for any particular wasp i, if its mother had been a P- foundress instead of a P+ foundress. We construct this counterfactual using the ratio of the headlengths for the offspring emerging from unpollinated figs versus pollinated figs. For any particular wasp i, then, the relative probability of becoming a foundress can be expressed as a function: $g(P(F|hl_i^-), P(F|hl_i^+))$. We use the ratio, so we define the *relative fitness* ($\mathbf{F_R}i$ as: $P(F|hl_i^-)/P(F|hl_i^+)$.

We then want to calculate the average relative fitness $\mathbf{F}_{\mathbf{R}i}$, averaged across the distribution of head lengths observed for that species. To do so, we impute a counterfactual head length for a wasp if its mother is a P-foundress, hl_i, by multiplying its observed head length hl_i^+ by the experimentally estimated ratio of head lengths for the daughters of P- and P+ wasps. Often this imputed head length hl_i was outside the lower range of the original data. Estimating $P(F|hl)$ for values of hl outside of the range of the original data is problematic since it would depend on how our density estimators extrapolate. In some cases, such as in *F. citrifolia* the counterfactual head length is far outside the lower range of the data. Rather than rely on parametric extrapolation, we conservatively opted to estimate a lower bound of the effect of being a daughter of a P- by rounding all imputed hl_i^- up to the minimum head length observed in the P+ data, where $P(F|hl)$ is non-parametrically identified. Denote this conservative imputed head length as hl_i^{-*}. For each emerging wasp we then calculate $\mathbf{F}_{\mathbf{R}i} = P(F|hl_i^{-*})/P(F|hl_i^+)$. We then average this ratio across the sample of emerging wasps: $\mathbf{F}_{\mathbf{R}} = \sum_i \mathbf{F}_{\mathbf{R}i}/N$. Thus, we report a conservative estimate of the average relative probability of becoming a foundress for P-offspring compared to P+ offspring, assuming that the P- head lengths are shifted by the experimentally estimated ratio for the P- offspring. In order to estimate a confidence interval for $\mathbf{F}_{\mathbf{R}}$ we redo the above procedure using each bootstrapped estimate of $P(F|hl)$.

1.3.3 Hypothesis test of equal probability of becoming foundresses of P- and P+ offspring

Finally, we formalize our comparison by conducting a hypothesis test of the sharp null that P- offspring have the same probability of becoming foundresses as P+ offspring because P- offspring have the same head length distributions as P+ offspring. To conduct this we perform the above, taking as our test statistic the above quantity: $\mathbf{F}_{\mathbf{R}}$. We bootstrap observations for both the emerging and foundress data from the combined emerging and foundress head length data, which are assumed under the null to be from the same distribution. Our alternative hypothesis is that P- offspring have a lower probability of becoming foundresses than P+ offspring (i.e. $\mathbf{F}_{\mathbf{R}} < 1$). Our p-value is then estimated by the proportion of bootstrapped samples that reveal an average ratio equal to or smaller than the observed ratio.

2 References

Blower, G., & Kelsall, J. E. (2002). Nonlinear kernel density estimation for binned data: convergence in entropy. *Bernoulli.* http://doi.org/10.2307/3318847.

Jandér, K. C. and E. A. Herre (2010). Host sanctions and pollinator cheating in the fig tree - fig wasp mutualism. *Proceedings of the Royal Society of London, B series* 277: 1481-1488.