Supporting Information. Disease where you dine: Plant species and floral traits associated with pathogen transmission in bumble bees. Lynn S. Adler, Kristen M. Michaud, Stephen P. Ellner, Scott H. McArt, Philip C. Stevenson, and Rebecca E. Irwin. *Ecology*. 2018.

Appendix S3. Statistical analysis details

Transmission trials across 14 plant species

Justification of analyzing transmission risk and intensity separately, instead of one response with negative binomial regression. The counts for each flower species were highly zero-inflated relative to a Poisson distribution with equal mean (16% to 53% zero counts, vs. <1% zero counts expected from Poisson distributions). Negative binomial regression is often used for such zero-inflated count data, but was not suitable for our data because the degree of zero-inflation varied substantially among species. In a negative binomial regression model, flower species with a higher intensity (i.e., higher mean of positive counts) would also have higher transmission risk (i.e., higher fraction of non-zero counts), but transmission and intensity were only weakly correlated (Pearson correlation coefficient $r = 0.39$, $p = 0.17$; Fig. S1). We therefore analyzed transmission risk and intensity as two separate components of pathogen transmission to bees.

Helianthus *as an outlier for floral traits and foraging behavior*. The distinctive floral architecture of *Helianthus* made it an outlier with respect to several floral traits, and resulted in very different foraging behavior. Bees on *Helianthus* probed over 400 disc flowers during a trial, more than twice the maximum number on any other species, and probed disc flowers and inoculum drops more than 5 and 10 times faster than the maximum risk of any other species, respectively. Several statistically significant apparent associations between traits and disease transmission were driven by a few trials with exceptionally active bees foraging on *Helianthus*. Our analyses of trait-dependent transmission therefore omitted *Helianthus*, but *Helianthus* was included in analyses that assessed species differences in transmission without considering floral traits.

Selecting final model for traits predicting intensity. Four variables (nectar production, corolla size, corolla shape, reproductive structures per inflorescence) were significant or marginally significant individual predictors in separate models predicting intensity (Table 2). Some of these variables are multicollinear (Pearson correlation $r = 0.93$ between nectar

production and corolla size, and r = -0.5 between each of these variables and reproductive structures per inflorescence), so a strong effect of one variable on intensity could give predictive power to the others when used as the sole predictor. To test for this possibility, we fitted linear models with reproductive structures per inflorescence and either corolla size, corolla shape, or nectar production as fixed-effect covariates; these models had less multicollinearity (VIF < 1.9). In all models, only reproductive structures per inflorescence was significantly related to intensity ($p < 0.03$ in all cases) while the other covariate was not ($p > 0.3$). The final model for intensity therefore had reproductive structures per inflorescence as the only fixed effect.

Evaluating bias in predictions of trait-based models. AIC evaluates only the magnitude of prediction errors. To assess whether predictions of traits-based models might be biased, we computed the predicted transmission risk and intensity for each trial using the final trait-based models. We averaged those predictions to obtain predicted mean transmission risk (Fig. S2A) and intensity (Fig. S2B) for each species, which can be compared to the observed mean transmission risk and intensity. Linear regressions through the plots of observed vs. predicted values (solid black line), were nearly identical to the 1:1 lines (dashed red line), so there is no evidence of bias in the traits-based predictions, either upward, downward, or towards the mean for all species.

Transmission trials manipulating flower number

Discarded data. Bees were discarded if they exhibited unusual foraging behavior (e. g., difficulty flying) or died prior to dissection. Two bees with abnormally high *Crithidia* counts (>200 cells/ 0.02 μL; one bee each from the *Lythrum* low flower and *Monarda* high flower treatments) were considered outliers and discarded prior to analysis. In total, 1, 2 and 5 bees were discarded from *Penstemon*, *Monarda* and *Lythrum* trials.

Adequacy of negative binomial model. The adequacy of the negative binomial model including all significant covariates for each species was tested by computing the Kolmogorov-Smirnov distance between the experimental data and the fitted negative binomial distributions for counts (using the R function ks.test), and doing the same for 500 artificial data sets simulated from the fitted distributions, identical in size and structure to the experimental data (generated by the simulate function for glm.nb fits). For each species, the K-S distance of the experimental data

was below the median K-S distance for the 500 artificial data sets, hence there is no evidence that the data for any species depart from the fitted negative binomial model.

Model selection. For *Penstemon*, no additional covariates predicted pathogen count ($p > 0.35$, $\chi^2 < 0.67$, $n = 65$ or 66). For *Lythrum,* trial time ($p = 0.035$, $\chi^2 = 4.46$, $n = 68$) and minutes to trial ($p = 0.044$, $\chi^2 = 4.061$, $n = 69$) were significant additional covariates. A second screening including those covariates as fixed effects and adding other covariates one at a time found that no other covariates were significant ($p > 0.3$, $\chi^2 < 0.88$, $n = 67$ or 68). The presence of a treatment effect was therefore tested in a model with trial time and minutes to trial. For *Monarda*, treatment ($p = 0.014$, $\chi^2 = 6.029$, $n = 51$) and number of flowers probed ($p = 0.002$, $\chi^2 = 9.841$, $n = 51$) were significant as predictors so a second screening was done with those as fixed effects and other covariates added one at a time; none of the other covariates were significant predictors ($p > 0.2$, $\chi^2 < 1.6$, $n = 50$ or 51). The effect of treatment was therefore tested in a model including flowers probed as a fixed effect.

Figure S1. Association between transmission and intensity across flower species. Species acronyms begin with the first 3 letters of the genus. Solid circles are at the point estimates of transmission and intensity for each species. Cross-species Pearson correlation between transmission and intensity was $r = 0.39$ ($p = 0.17$); with *Helianthus* removed this becomes $r = 0.35$ ($p = 0.23$). Source file: SusceptibilityAndIntensityPlots.R

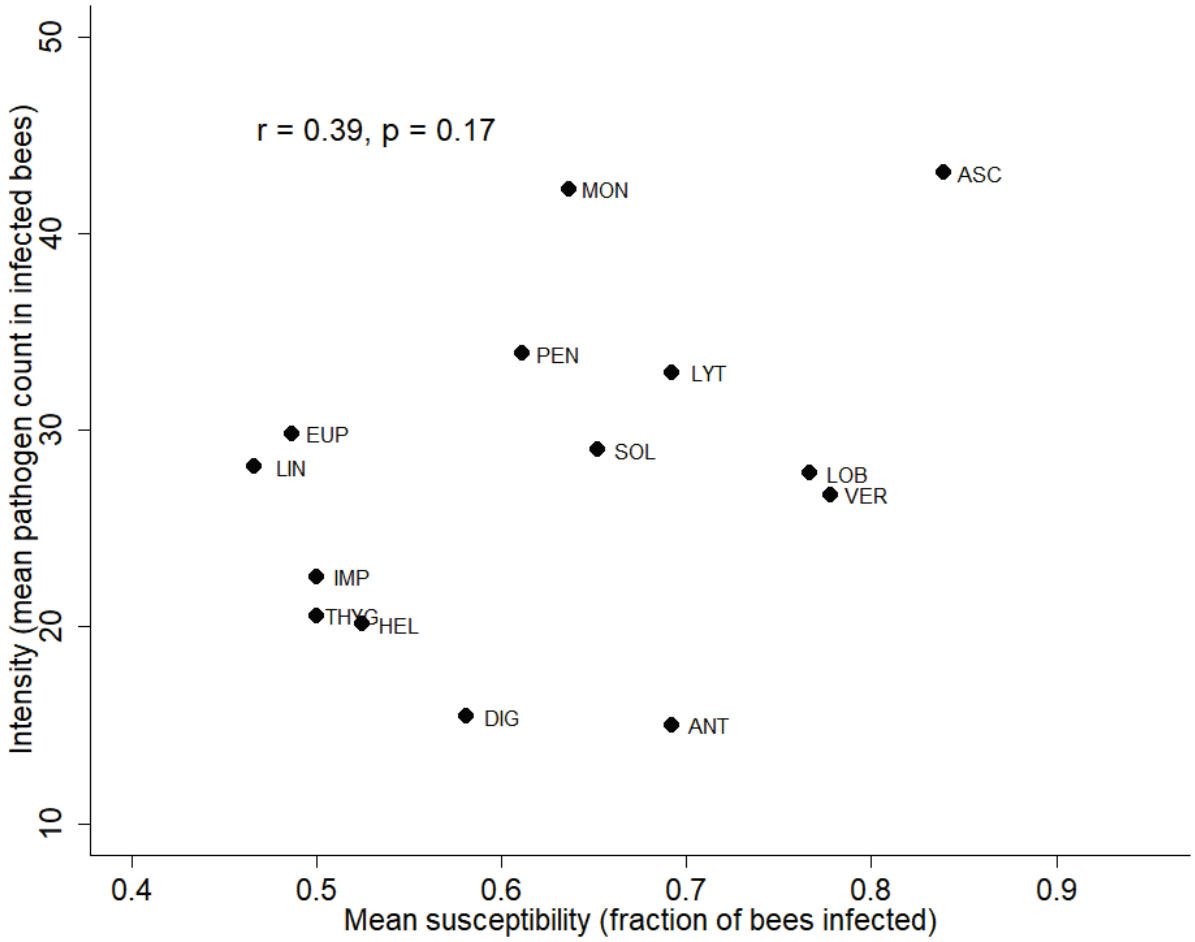

Fig. S2. Plots of A) observed mean susceptibility and B) observed mean intensity, for each species, versus predictions from the final traits-based model. In each panel, the solid black line is the regression line fitted to the plotted species-specific values (for observed and predicted transmission risk and mean intensity, respectively, in panels A and B) and the dashed red line is the 1:1 line). The regression line would coincide with the 1:1 line if predictions are unbiased. Source files: SpeciesTraitsAndSusceptibility.R SpeciesTraitsAndIntensity.R